Calibration of Low Cost Sensor using Machine Learning Techniques

B.Tech. Project Report

by

Swadesh Choudhary

1906338

Faculty Adviser

Dr. Thaseem Thajudeen

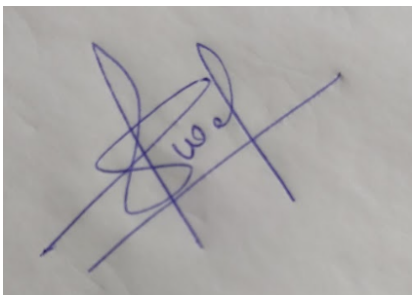
School of Mechanical Sciences

Indian Institute of Technology Goa

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources.

I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.



(Signature)

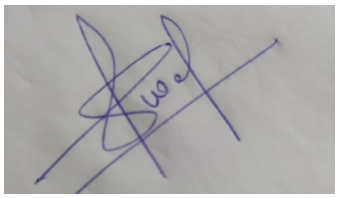
Swadesh Choudhary

(1906338)

Date: 10/4/2023

Abstract

The recent findings in air quality sensors have permitted the government and other communities to measure air pollution with higher temporal and spatial resolution. The use of low-cost sensors (LCS) for monitoring pollutants has increased drastically in the last few years, but there are missing links in the design and large-scale deployment of LCS. The challenges in LCS are analyzing the performance of sensors, developing a reliable method to deploy sensors, and communicating the sensor data in a meaningful way to the public. These sensors can be a good addition for different monitoring sites and provide high spatial-temporal PM mapping. The objective of this project was to do calibration to increase the correlation between the LCS (a Plantower PMS A003, PMS 5003 and Winsen) and the Reference Monitor Sensor (Alphasense for Particulate Matters (PM)) and to decrease the Mean Absolute error (MAE) and Root Mean Square Error (RMSE) using machine learning algorithms like Linear Regression, Multiple Regression, K-Nearest Neighbour, Random Forest, Gaussian Algorithm, Support Vector Method and Artificial Neural Network. The LCS are highly sensitive towards environmental factors that's why we considered all the factors like relative humidity, temperature, wind speed and wind direction. PMS 5003 is showing best prediction performance with random forest algorithm, support vector algorithm and gaussian process algorithm with the pearson correlation coefficient (r) of 0.853, 0.851 and 0.85. This study also investigates the performance of low cost sensors in the characterisation chamber by comparing it to Dusttrak 8533. The correlation between them has been analyzed by varying the humidity. At lower PM concentrations the value of LCS and Reference monitor are nearly close to each other but at higher PM concentrations the value of LCS deviates with respect to reference monitor.



Swadesh Choudhary

11.04.2023

Dr. Thaseem Thajudeen

Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Introduction

The World Health Organization (WHO) recently announced the revised recommendations for air-quality guidelines to urge authorities worldwide to effectively reduce air pollution for the protection of human health (Lung et al. 2022). According to estimates, 2.2 million of the 7 million premature deaths worldwide each year that are linked to household (indoor) and ambient (outdoor) air pollution deaths from both acute and chronic health impacts occurred in Asia. The major sources of air pollution are power plants, transportation, burning of fossil fuels, constructions, industries etc. which produce various harmful pollutants like Carbon dioxide (CO_2), Carbon monoxide(CO), Particulate matter(PM) and these pollutants are directly linked to the health of humans. Out of which the PM particle is one of the major problems of the world.

Urban areas frequently have significant amounts of PM pollution from both manmade and natural causes. Forest fires, dust storms, and volcanoes are the primary causes of natural PM emissions. One of the main causes of anthropogenic PM2.5 is traffic and vehicle emissions. Residential areas close to busy highways are also exposed to higher levels of PM. Other anthropogenic PM emissions come from the combustion of related n crustal materials, industrial processes, building sites, and mining.

The impact of fine particulate matter (PM2.5) on public health is a major topic of discussion. The fifth-ranking risk factor for death according to the Global Burden of Disease (GBD) Study 2015 (GBD 2015 Risk Factors Collaborators, 2016) was PM2.5, which was responsible for 4.2 million deaths and 103.1 million fewer disability-adjusted life-years (DALYs), or 7.6% of all deaths and 4.2% of all DALYs worldwide. According to epidemiological research, breathing in PM2.5 increases the risk of developing and dying from respiratory illnesses. The vulnerability to respiratory system infection may be attributed to the host defense failure brought on by PM2.5 exposures.

Particulate matter are solid and liquid particles which are suspended in air. The average diameter of hair is 60μm which is approximately 6 times the size of PM_{10} particles because of that it is difficult to measure up until now. Rapidly developed sensor technology has shed light on tackling this challenge. The depth to which the particles pierce the respiratory tract and their ability to harm cells depends on their size. Therefore, according to the size of Particulate Matters, it broadly classified into 3 categories:

- Coarse (PM_{10}) = Whose size is less than 10µm are considered as PM_{10} particles. When we inhale the air, generally these PM particles with an aerodynamic diameter 2.5–10 µm get stuck in our trachea and can easily be removed by drinking water and gargling.
- Fine ($PM_{2.5}$) = Whose size is less than 2.5µm are considered as $PM_{2.5}$ particles. PM less than 2.5 µm in diameter poses the greatest problems, because it can get deep into the terminal bronchioles and alveoli.
- Ultra Fine (PM_1) = Whose size is less than 1µm are considered as PM_1 particles. PM particles with size less than 1µm in diameter may even get into the bloodstream affecting other organs.

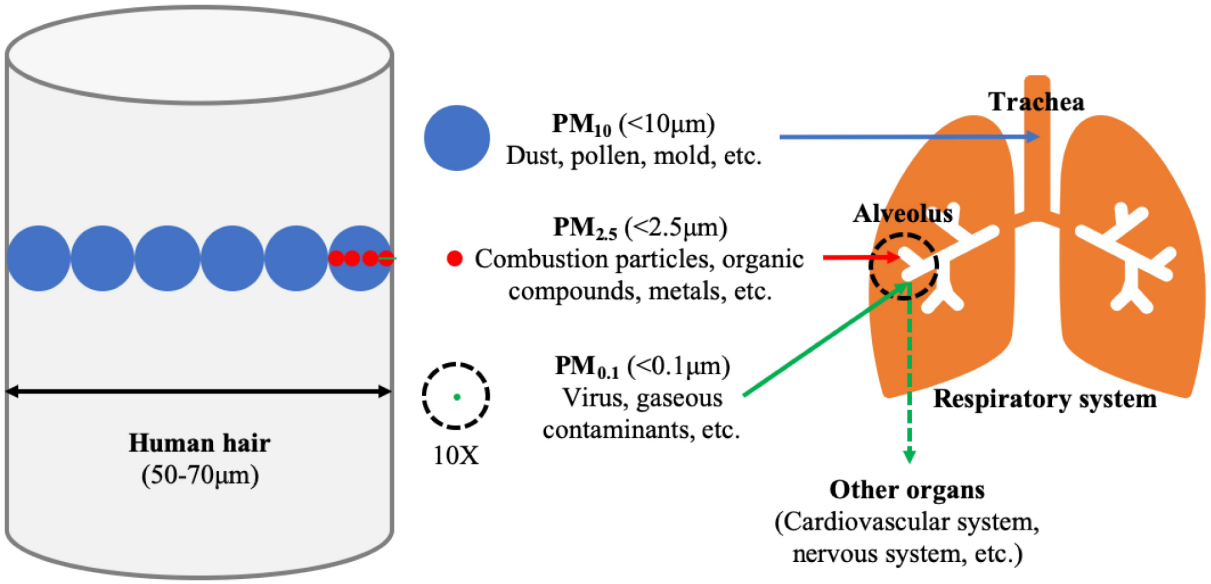


Figure 1.1 The size, main composition and deposition site in the lung of the particulate matter (PM)

Personal exposure to pollutants depends on the pollutant concentration as well as the contact time involved (Emily G Snyder et al. 2013). The concentration of air pollutants varies with space and time, due to which conventional air monitoring systems like gravimetric method and beta attenuation method (BAM) cannot measure the actual concentration to which people are exposed in the microenvironment. Therefore, there is a shift from conventional to low cost sensor (LCS) based air quality measurement, particularly for monitoring PM concentrations (Fadhli et al. 2022). The PMS 5003, PMS 7003, Winsen ZH 06, Sensirion SPS30, etc., are some commonly used LCS based on the light scattering measurement technique which is discussed in the next chapter. The objective of my research

is to calibrate the LCS with some reference monitor. The Low Cost Sensor(LCS) calibration will be done to increase the correlation between the LCS (PMS A003, 5003 and Winsen) and reference monitors (Alpha sense OPC N3), and Machine Learning techniques will be used for improved calibration.

1. Multiple studies in the recent past have considered factors like relative humidity, temperature, and particle size therefore we use the ML algorithm considering all the factors like relative humidity, temperature, and wind speed.
2. The Artificial Neural Network, Linear Regression, Multiple Linear Regression, K-Nearest Neighbor, Random Forest and Support Vector Method techniques will be used for correlation.
3. To mitigate the effect of relative humidity, an air drying system or heater will be used at the inlet of low cost sensors.
4. Effect on correlation from relative humidity and temperature individually will also be seen.

Typically, the performance of PM sensors is assessed inside of an aerosol chamber where physical parameters may be maintained and to eliminate the environmental disturbance. (Wang et al., 2015; Manikonda et al., 2016; Li and Biswas, 2017). An acrylic sheet of 68×60×52 cm^3 dimension is used to make a characterization chamber for the particle measurement in which aerosol particles will be added after, and the PM concentration will be monitored until it returns to normal and we have also seen the influence of humidity on PM concentration and variation in correlation at different point inside the characterization chamber when aerosol particles are inserted.

Chapter 2, discusses the description of different methods for air monitoring to measure particle size and concentration. Furthermore, the discussion on different conventional monitoring instruments, dynamic air monitors, and principles and working of LCS is explained. In addition, the different issues that can be challenging during the measurements are discussed. In chapter 3, data calibration and distribution of dataset into training and testing set. In Chapter 4, Numerical analysis approach used to calibrate the data obtained by different low cost sensors. In Chapter 5, the results and detailed analysis of experiments conducted inside the characterization chamber. In chapter 6, the results and detailed analysis of the performance of LCS at different locations in the Indian Institute of Technology Goa are discussed and the conclusions and plans for future studies are provided.

Chapter 2: Literature Review

According to the World Health Organization (WHO), air pollution accounts for more than 3.8 million deaths annually and is one of the leading causes of sickness and early mortality in the developing world (Kuula et al., 2022). Along with considerable technological advancements that LCS, the public's awareness of air quality issues has increased. The needs of citizens who desire online, real-time information on air quality as a part of their digital ecosystem can be met by these sensor systems. Due to the LCS's ability to produce high-density data based on geographical and temporal variation, their rapid diffusion over the past few years has been driven (Bennett et al., 2019).

The size of the PM particle is very small because of that it is difficult to measure up until now. Rapidly developed sensor technology has shed light on tackling this challenge. To monitor the mass concentration of PM particles some advanced technologies are used in real time. They provide continuous measurements of particle concentrations, producing valuable data and enabling trends to be tracked. If PM levels exceed a specified threshold, ventilation may be raised or an alert for changing the air filters may be set off. This chapter provides a comprehensive discussion of the latest updates on research areas related to low-cost sensors, the setting up of sensor-based networks, and an understanding of the capabilities of LCS

2.1 Conventional air monitoring system

Conventional grade ambient air monitors and instruments for pollutant characterization have significantly improved over the years, and the use of low-cost monitors is also expanding, which is being added to the arsenal of tools for air monitoring (Kumar et al. 2015). The typical measurements used by conventional air monitoring instruments include gravimetric techniques, cascade impactors, and tapered elementary oscillatory microbalances (TEOM), which are expensive and cumbersome (Kan et al. 2012). Due to the high expense of using conventional monitoring systems to obtain improved temporal resolution, dynamic air monitoring and inexpensive monitoring technologies have become essential.

2.2 Emergent or low-cost air monitoring system

The monitoring of air with low-cost PM sensors has become a significant choice nowadays because of its low cost, compact size, adequate accuracy, minimum power

requirement, etc. LCS deployment has its own challenges, including the need for continuous power supply and network requirements.

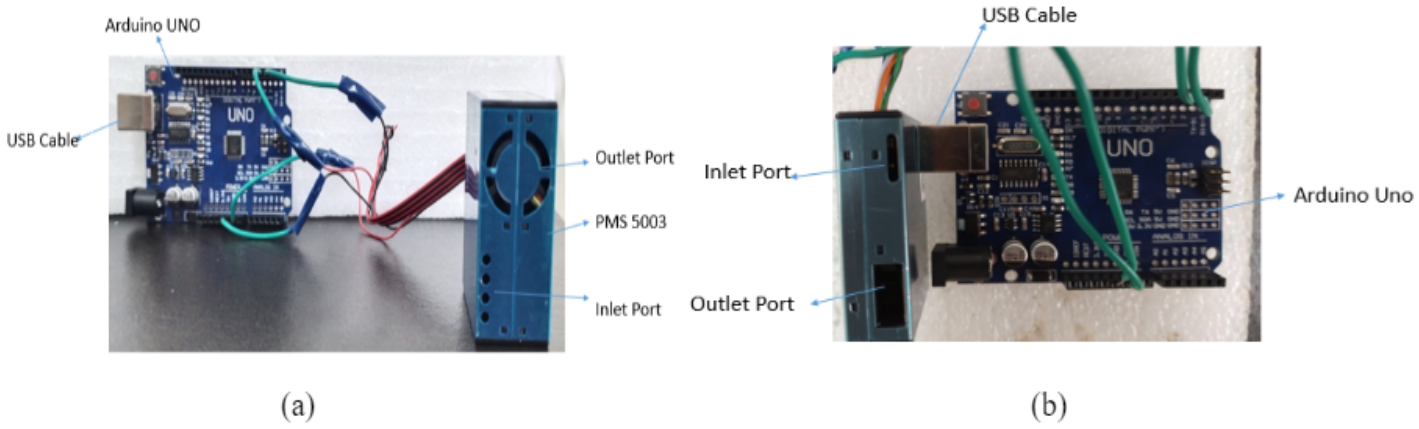


Figure 2.1 Low cost sensor (a) PMS 5003 and (b) Winsen

Table 2.1 Comparison between Reference Sensors and Low Cost Sensors

Reference Sensors	**Low cost Sensors**
Highly Accurate	Less Accurate
Bulky	Manageable
Required high maintenance	low maintenance required
Not easily affordable because of its cost	Easily affordable
BAM, TEOM, Dusttrak	PMS A003, PMS 5003, Winsen

2.2.1 Working of low cost sensor

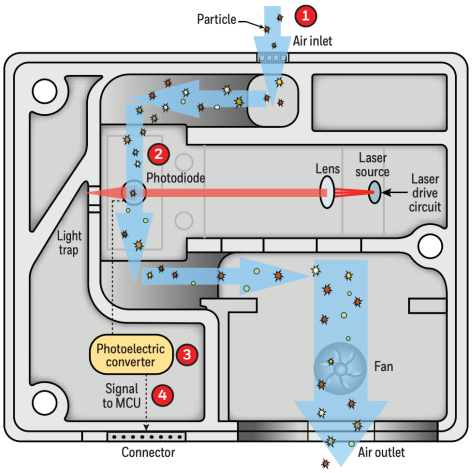


Figure 2.2 Functional diagram of PM Sensor

The PM sensor uses a laser scattering principle, which works by using a laser to radiate suspending particles in the air. Then scattering light is collected to a certain degree, and finally the curve of scattering light changes with time. Nephelometric measures the particle size, and the principle of this device is when the particles flow through the cavity, the light strikes the particles and gets modulated before reaching the phototransistor. Nephelometric response intensity is calculated, and with the help of some relation with intensity, the mass concentration is calculated (Williams et al., 2019). When the light beam hits the particle, it undergoes scattering. Light scattering can also be defined as light dispersion by a particle and use a photometer detector to measure the scattered light intensity at many angles. Multiple studies describe the light scattering photometers used to detect the scattered light from the particles in the detection zone.

2.2 Characterization Chamber

The creation of LCS involves a lot of work in order to ensure data reliability. The sensors must be thoroughly characterized for various environmental conditions and calibrated using reference devices. In order to facilitate the various environmental testing conditions, testing chambers are required. Before being deployed at various outdoor or indoor measurements, different LCS can have their performance evaluated by being tested in a characterization chamber with varying aerosol concentrations and ambient factors. The laboratory tests demonstrate the behavior of LCS in various particle compositions and environmental settings. Recent studies that examined direct in-field calibration without laboratory evaluation discovered good agreement between the sensor and accepted practices (Moreno-Rangel et al. 2018). One of the first test chambers was created by (Wang et al. 2015) and assessed the LCS performance in a chamber with uniformly distributed particles of various concentrations. The same year, (Austin et al. 2015) created an airtight testing chamber whose volume was reduced by using baffles in addition to the box and was 6x12x8 cm3. However, it can be inferred from various studies (He et al. 2020; Chatzidiakou et al. 2019; Hapidin et al. 2019) that laboratory calibration does not fully capture all of the characteristics of a specific location where sensors will be placed, making it insufficient to predict how sensors will behave in real-time. In the next chapter we will discuss different steps that are followed for data analysis and data training. The steps and results from the characterization chamber are also analyzed.

Chapter 3: Data Calibration

3.1 Data Preparation

All commercially available low-cost PM sensors work on the basis of light scattering. In comparison to the majority of other techniques for particle counting and mass concentration measurements, light scattering can be used in much smaller form factors. In theory, all that is needed for a low-cost PM sensor is a light-emitting diode (often an infrared or red laser diode), a phototransistor, and a lens to concentrate the diode light.

The sensors are placed at IIT Goa campus and following steps were taken to process the raw data. The interval data recorded by the low-cost sensor, including PM2.5, temperature, and RH, were averaged into hourly data because the sensor is calculating the data every minute therefore to reduce the noise we did the averaging of the data.

3.2 Low-cost sensor evaluation

Data calibration must be carried out. To do this, various regression algorithms will be trained against the reference data (collected by Alpha Sensor) and the data collected by the network of Low Cost Sensor (LCS, Temp, and RelHum, the explanatory variables) and we have also shown the individual effect of all the sensors in algorithms.

We have collected the data from three different Low Cost Sensors(LCS) that are PMS A003, PMS 5003 and Winsen for $PM_{2.5}$

The primary dataset is divided into two datasets (train and test) for this purpose, with a split of 75% to 25%. While the test dataset has a test size of 0.25, the train dataset has a train size of 0.75. The dataset includes information for three consecutive months.

The algorithms are trained against the dataset of 3 months and then the result is calculated on the basis of the test dataset.

To calculate the performance of our low cost sensor we have also calculated the Loss function:

- Root-mean-square error (RMSE): The standard deviation of the residuals (prediction errors). It indicates how close the observed data points are to the model's predicted values.
- Mean absolute error (MAE): It measures the errors between paired observations expressing the same phenomenon.
- Pearson correlation coefficient (r): It is a statistical measure that indicates the extent to which two or more variables fluctuate in relation to each other.

3.3 Dataset

The dataset consists of the data of Particulate Matter(PM) concentration collected by a low cost sensor (LCS) and the reference sensor(Alpha Sensor) at IIT Goa. It contains the data of 3 months on which calibration is done.

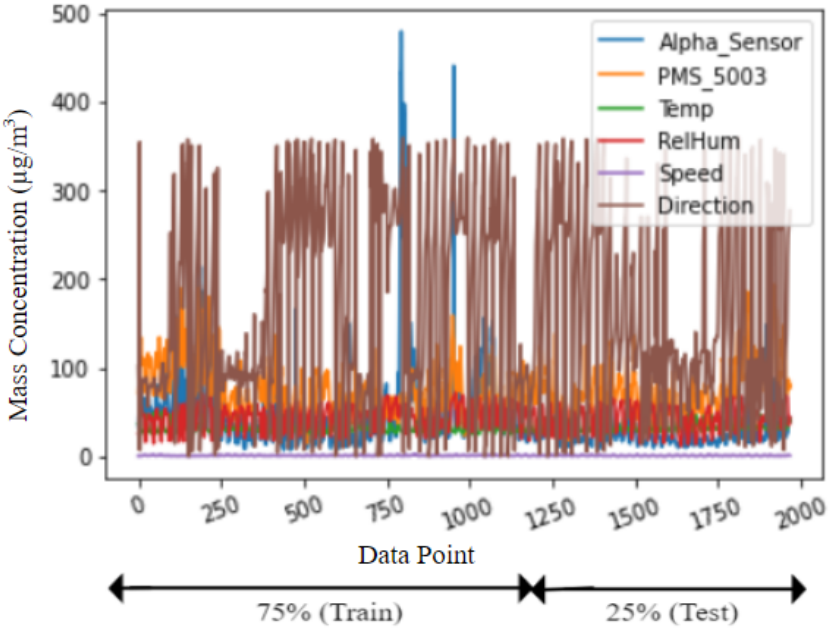


Figure 3.3 Variation of different sensors with respect to data points

The data is organized as follows:

1. Alpha Sensor: Reference sensor, in µg/m^3 (real PM concentration levels)
2. Sensor: Low cost sensor, in µg/m^3 (inaccurate PM concentration, to be calibrated)
3. Temp: Temperature sensor, in °C
4. RelHum: Relative humidity sensor, in %
5. Speed: Wind Speed, in m/s
6. Direction: Wind Direction, in degrees

In the next chapter we will discuss the different machine learning algorithms and how to train the data set from these algorithms. We will also see the effect of relative humidity and temperature on PM concentration individually.

Chapter 4: Numerical Analysis

This chapter contains the air sampling data collected for the different PM particles concentration from different sensors from December 2021 to March 2022. The data obtained from the different sensors are calibrated with alpha sense OPC N3 by using different machine learning techniques.

4.1 Linear Regression

Linear regression is a simple and important method for predictions in the analysis. In Linear Regression, a relationship is established between independent and dependent variables by fitting them to a line. This line is known as the regression line and is represented by a linear equation.

$$Y= a *X + b$$

where in this equation:

- Y – Dependent Variable
- a – Slope
- X – Independent variable
- b – Intercept

In our case the independent variable is the data obtained from the low cost sensor(LCS) and the dependent variable is the data obtained from the reference sensor(Alpha Sensor).

4.1.1 Results from linear regression

By using linear regression, we obtained following results for different low cost sensor for $PM_{2.5}$ as:

The graph below compares the real concentration levels at the reference sensor (Alpha sensor) with the obtained predicted values by low cost sensor (PMS 5003, PMS A003, Winsen) after calibration with Linear Regression

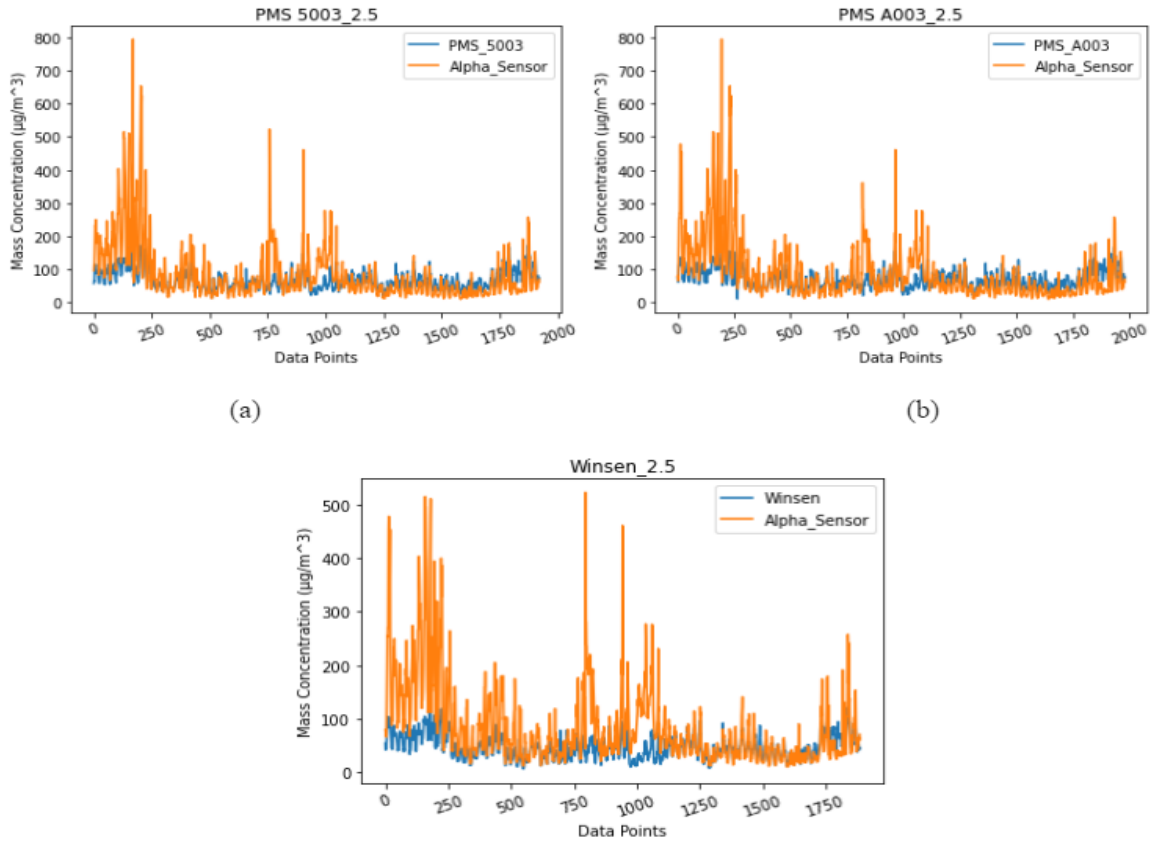


Figure 4.1 Relation between (a) PMS 5003 and alpha sensor, (b) PMS A003 and alpha sensor (c) Winsen and alpha sensor obtained from linear regression

Table 4.1 Obtained results from Linear regression when both relative humidity and temperature effect considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	65.15	62	38.67
Mean Absolute Error (MAE)	61.34	57.7	35.31
Correlation	0.62	0.67	0.68

Table 4.2 Obtained results from Linear regression when only temperature is considered

Loss Function	PMS A003	PMS 5003	Winsen

Root Mean Square Error (RMSE)	65.15	62	38.67
Mean Absolute Error (MAE)	61.34	57.7	35.31
Correlation	0.62	0.67	0.68

Table 4.3 Obtained results from Linear regression when only relative humidity is considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	65.15	62	38.67
Mean Absolute Error (MAE)	61.34	57.7	35.31
Correlation	0.62	0.67	0.68

4.2 Multiple Linear Regression

Multiple Regression is the extension of linear regression, In which a relationship is established between independent and dependent variables by fitting them to a line. However, in this we have more than one independent variable. Therefore, in this we have considered all the necessary parameters like Relative Humidity and Temperature to obtain correlation.

$$y = a_0x_0 + a_1x_1 + a_2x_2 + a_3x_3 + \ldots + a_nx_n + b$$

where in this equation:

- x_0, x_1, x_2, x_3, x_n – Dependent Variable
- a_0, a_1, a_2, a_3, a_n – Coefficient of dependent variable
- y – Independent variable
- b – Intercept at y axis

In our case the independent variable is the data obtained from the low cost sensor(LCS), Relative Humidity, Temperature and the dependent variable is the data obtained from the reference sensor(Alpha_Sensor).

To obtain the values of coefficient of dependent variable and intercept we used sklearn's fit function.

$$\text{Pred} = \beta_0 + \beta_1 \cdot (\text{Low Cost Sensor}) + \beta_2 \cdot \text{Temp} + \beta_3 \cdot \text{RelHum}$$

where in this equation:

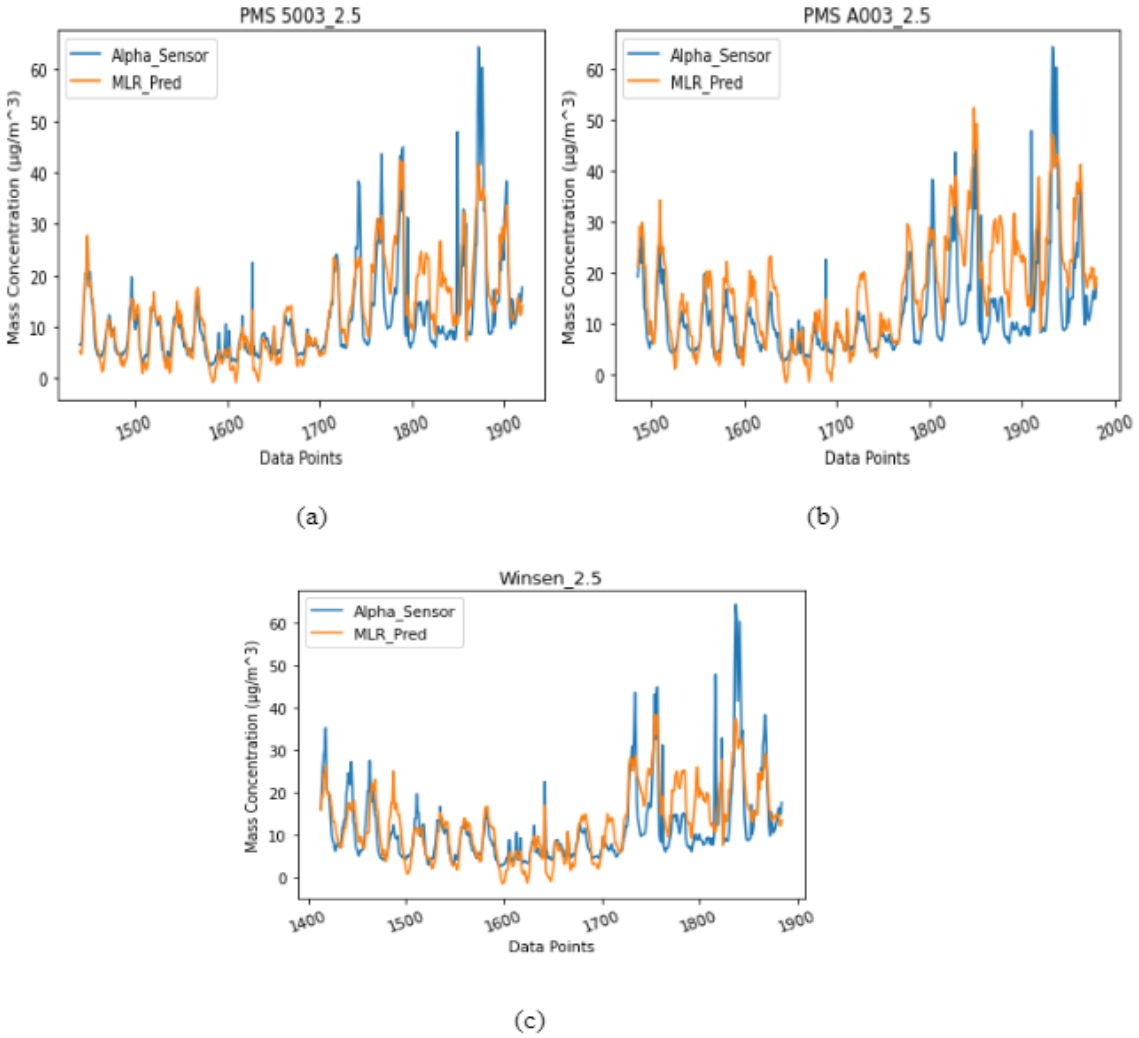
- Pred = Value that we are going to predict using machine learning algorithms and training the dataset.
- $\beta_0, \beta_1, \beta_2, \beta_3$ = Constants values corresponding to input parameters
- Low Cost Sensor = Data recorded by our low cost sensors.
- Temp: Temperature sensor, in °C
- RelHum: Relative humidity sensor, in %

Table 4.4 Values of constraint terms calculated by using sklearn' s fit function

	Low Cost Sensor		
Constant term	PMS A003	PMS 5003	Winsen
β_o	22.8	24.02	23.24
β_1	14.98	14.63	14.26
β_2	0.82	1.66	1.65
β_3	2.27	4.74	3.83

4.2.1 Results from Multiple Linear Regression

The graph below compares the real concentration levels at the reference sensor (Alpha_sensor) with the obtained predicted values by low cost sensor (PMS 5003, PMS A003, Winsen) after calibration with Multiple Linear Regression.



4.2 Relation between (a) PMS 5003 and alpha sensor, (b) PMS A003 and alpha sensor (c) Winsen and alpha sensor obtained from Multiple linear regression

Table 4.5 Obtained results from Multiple Linear regression when both relative humidity and temperature effect considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	18.37	17.53	5.85
Mean Absolute Error (MAE)	15.24	13.9	4.16
Correlation	0.83	0.849	0.787

Table 4.6 Obtained results from Multiple Linear regression when only temperature is considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	18.37	17.53	5.85
Mean Absolute Error (MAE)	15.24	13.9	4.16
Correlation	0.83	0.849	0.787

Table 4.7 Obtained results from Multiple Linear regression when only relative humidity is considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	18.37	17.53	5.85
Mean Absolute Error (MAE)	15.24	13.9	4.16
Correlation	0.83	0.849	0.787

4.3 K-Nearest Neighbor

KNN is an easy-to-use algorithm that sorts new cases by getting the consent of the majority of its k neighbors before storing them altogether. The class with which the case shares the most characteristics is then given a case. This calculation is made using a distance function.

Step-1: Select the number K of the neighbors

Step-2: Calculate the Euclidean distance of K number of neighbors

Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.

Step-4: Among these k neighbors, count the number of the data points in each category.

Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.

Step-6: Our model is ready.

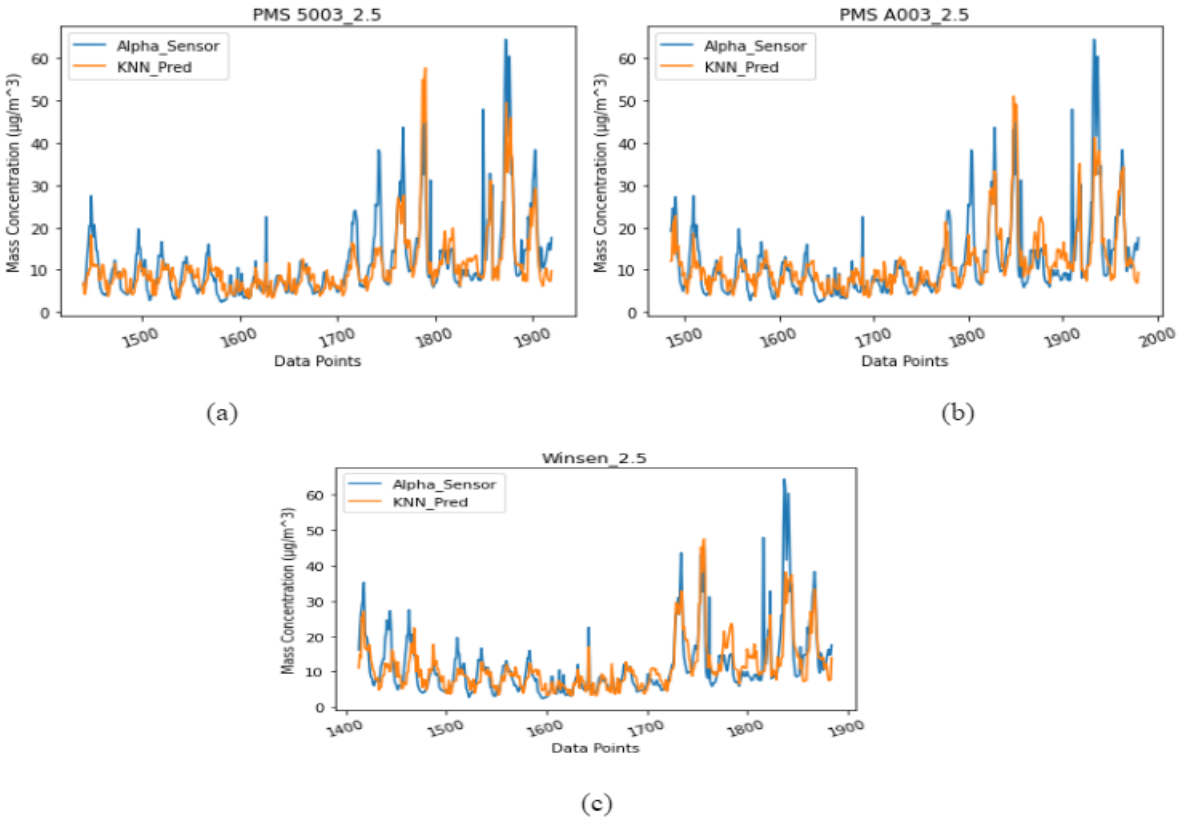
where;

$$\text{Euclidean distance (d(x,y))} = \sqrt{\sum_{i=1}^{n} (Yi - Xi)^2}$$

As K-Nearest Neighbor works on the value of K which is number of nearest neighbor, now taking best possible value of K is very important because if we consider less value than the data would be under fitted similarly if we choose large value of K then the data would be over fitted Therefore to get best possible value of K we obtained the results by changing the values of K from 1 to 50 and we get best result when K = 19.

4.3.1 Results from K-Nearest Neighbor

The graph below compares the real concentration levels at the reference sensor (Alpha sensor) with the obtained predicted values by low cost sensor (PMS 5003, PMS A003, Winsen) after calibration with K-Nearest Neighbor.



4.3 Relation between (a) PMS 5003 and alpha sensor, (b) PMS A003 and alpha sensor (c) Winsen and alpha sensor obtained from K-Nearest neighbor algorithm

Table 4.8 Obtained results from K-Nearest neighbor algorithm when both relative humidity and temperature effect considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.32	5.09	5.41
Mean Absolute Error (MAE)	3.63	3.31	3.56
Correlation	0.817	0.841	0.811

Table 4.9 Obtained results from K-Nearest neighbor algorithm when only temperature effect considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.42	5.21	5.49
Mean Absolute Error (MAE)	3.71	3.45	3.91
Correlation	0.816	0.833	0.807

Table 4.10 Obtained results from K-Nearest neighbor algorithm when only relative humidity effect considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.42	5.21	5.49
Mean Absolute Error (MAE)	3.71	3.45	3.91
Correlation	0.816	0.833	0.807

4.4 Random Forest

The term "Random Forest" refers to a collection of decision trees. Each tree is assigned a class, and the tree "votes" for that class, in order to categorize a new item based on its characteristics. The classification with the highest votes is selected by the forest (over all the trees in the forest).

In a regression problem, a random forest algorithm provides the mean or average forecast made by each individual tree. Random forests are made up of trees that grow in parallel with no contact between them.

In Random Forest, there is just one hyperparameter that needs to be set: the number of trees (n estimators). RMSE, MAE, Correlation, and time to solve are performance statistics that are calculated in order to fine-tune this parameter for this issue (in ms). These variables are shown against the n estimators value, which spans from 1 to 100 estimators. From the result we have seen that we are getting the best result when the value of n estimators is 20.

4.4.1 Results from Random Forest

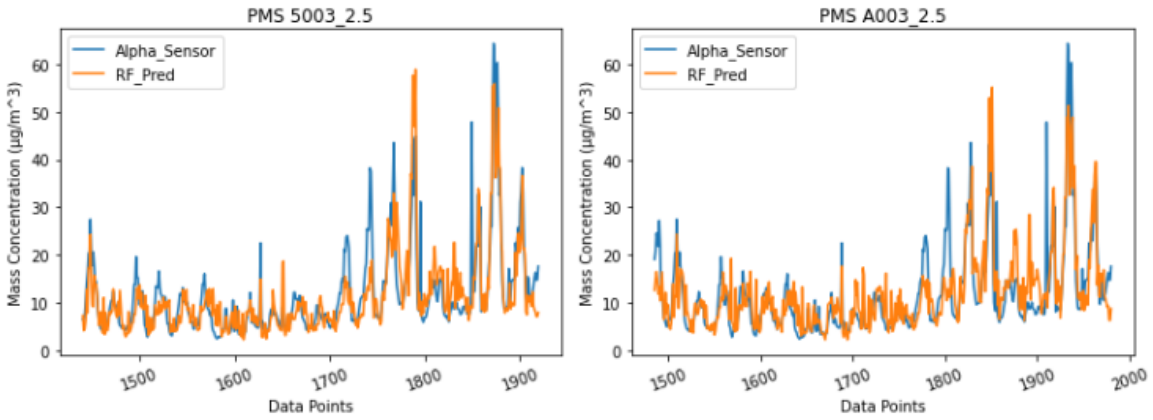
The graph below compares the real concentration levels at the reference sensor (Alpha_sensor) with the obtained predicted values by low cost sensor (PMS 5003, PMS A003, Winsen) after calibration with Random Forest

Table 4.11 Obtained results from random forest algorithm when both relative humidity and temperature effect considered

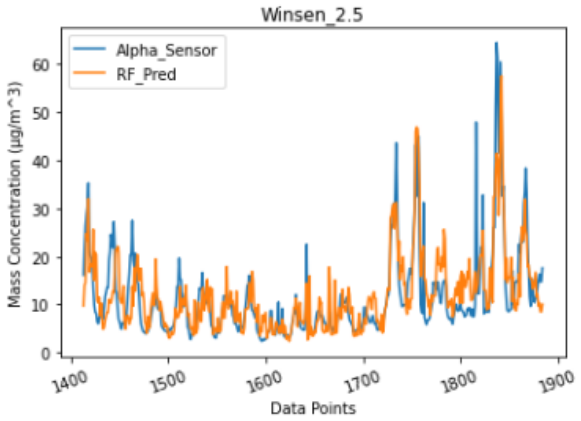
Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.42	5.21	5.49
Mean Absolute Error (MAE)	3.71	3.45	3.91
Correlation	0.816	0.833	0.807

Table 4.12 Obtained results from random forest algorithm when only temperature effect is considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.42	5.21	5.49
Mean Absolute Error (MAE)	3.71	3.45	3.91
Correlation	0.816	0.833	0.807



(a) (b)



(c)

4.4 Relation between (a) PMS 5003 and alpha sensor, (b) PMS A003 and alpha sensor (c) Winsen and alpha sensor obtained from random forest algorithm

Table 4.13 Obtained results from random forest algorithm when only relative humidity effect is considered

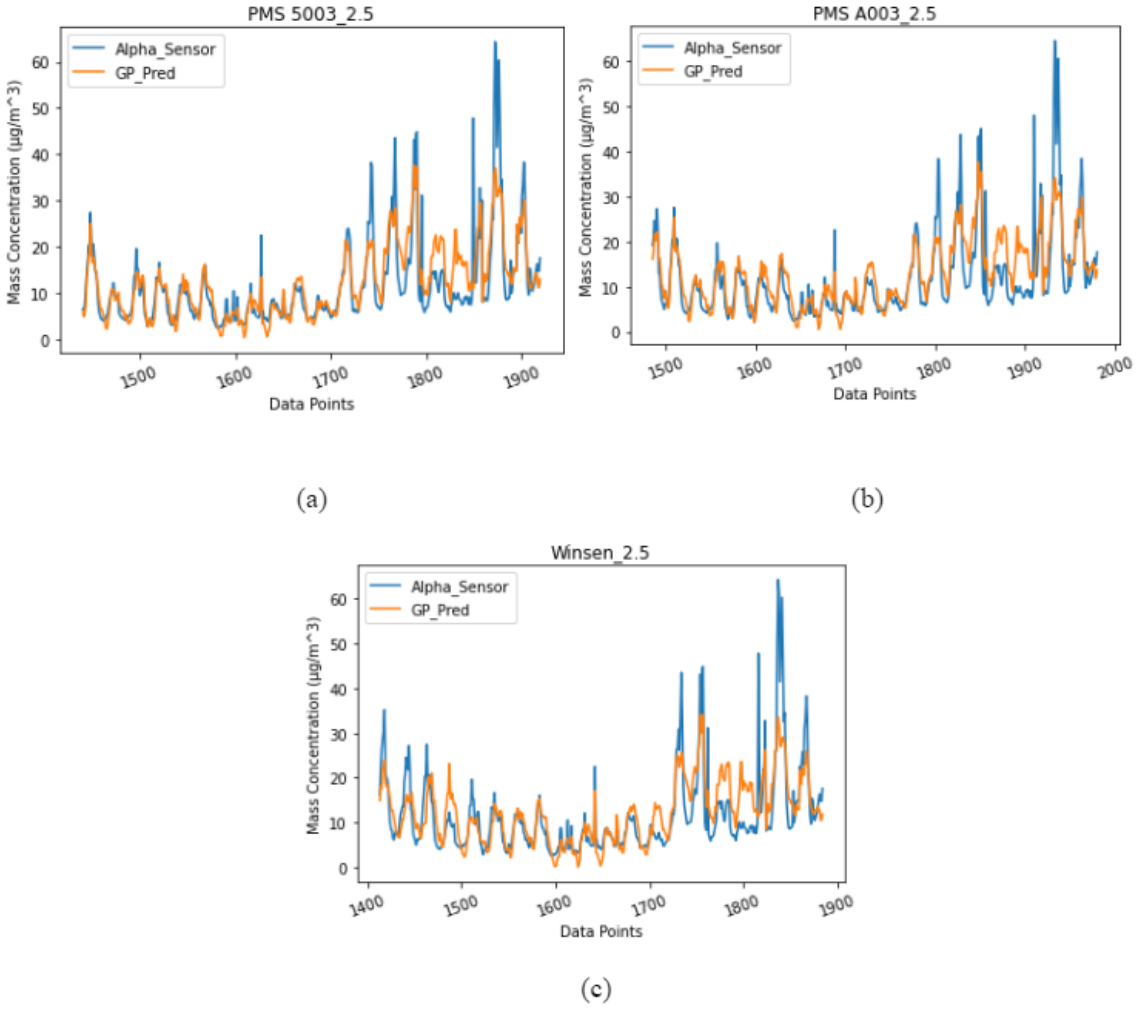
Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.42	5.21	5.49
Mean Absolute Error (MAE)	3.71	3.45	3.91
Correlation	0.816	0.833	0.807

4.5 Gaussian Progression

Gaussian Process Regression (GPR) is a remarkably powerful class of machine learning algorithms, it relies on few parameters to make predictions. From the given dataset it Creates all the possible probability distribution functions and Based on our probability measure a reasonable guess is the mean of the all function.

4.5.1 Results from Gaussian Progress

The graph below compares the real concentration levels at the reference sensor (Alpha_sensor) with the obtained predicted values by low cost sensor (PM 5003) after calibration with Gaussian Process



4.5 Relation between (a) PMS 5003 and alpha sensor, (b) PMS A003 and alpha sensor (c) Winsen and alpha sensor obtained from Gaussian process algorithm

Table 4.14 Obtained results from Gaussian process algorithm when both relative humidity and temperature effect considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.32	4.93	5.74
Mean Absolute Error (MAE)	3.56	3.08	3.86
Correlation	0.83	0.848	0.785

Table 4.15 Obtained results from Gaussian process algorithm when only temperature effect is considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.32	4.93	5.74
Mean Absolute Error (MAE)	3.56	3.08	3.86
Correlation	0.83	0.848	0.785

Table 4.16 Obtained results from Gaussian process algorithm when only relative humidity effect is considered

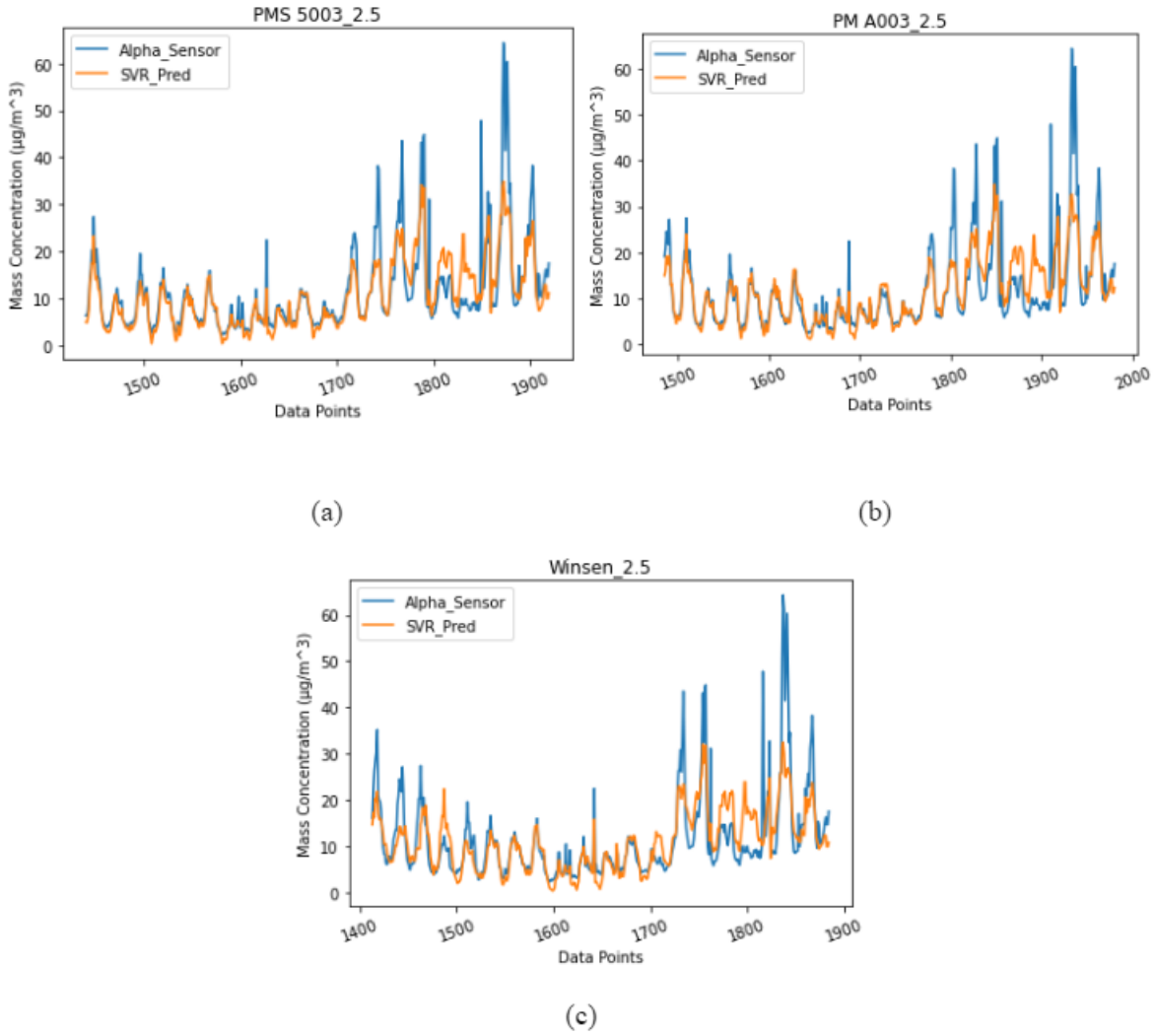
Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.32	4.93	5.74
Mean Absolute Error (MAE)	3.56	3.08	3.86
Correlation	0.83	0.848	0.785

4.6 Support Vector Method

SVM algorithm is a method of a classification algorithm in which you plot raw data as points in an n-dimensional space (where n is the number of features you have). The value of each feature is then tied to a particular coordinate, making it easy to classify the data. The data can be divided into groups and plotted on a graph using lines known as classifiers.

4.6.1 Results from Support Vector Method

The graph below compares the real concentration levels at the reference sensor (Alpha_sensor) with the obtained predicted values by low cost sensor (PM 5003) after calibration with Support Vector Method



4.6 Relation between (a) PMS 5003 and alpha sensor, (b) PMS A003 and alpha sensor (c) Winsen and alpha sensor obtained from support vector method algorithm

Table 4.17 Obtained results from support vector method algorithm when only relative humidity effect is considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.2	5.07	5.74

Mean Absolute Error (MAE)	2.9	2.83	3.54
Correlation	0.83	0.846	0.788

Table 4.18 Obtained results from support vector method algorithm when only temperature effect is considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.2	5.07	5.74
Mean Absolute Error (MAE)	2.9	2.83	3.54
Correlation	0.83	0.846	0.788

Table 4.19 Obtained results from support vector method algorithm when only relative humidity effect is considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.2	5.07	5.74
Mean Absolute Error (MAE)	2.9	2.83	3.54
Correlation	0.83	0.846	0.788

4.7 Artificial Neural Network

A neural network is a method in artificial intelligence that is inspired by the human brain. It includes input layer, output (or target) layer and, in between, a hidden layer. The layers are connected via nodes, and these connections form a “network” – the neural network. It is widely used in forecasting, marketing research, fraud detection and risk assessment.

In our case the Input layers are relative humidity, temperature, low cost sensors obtained value and the output layers are the calibrated $PM_{2.5}$ values.

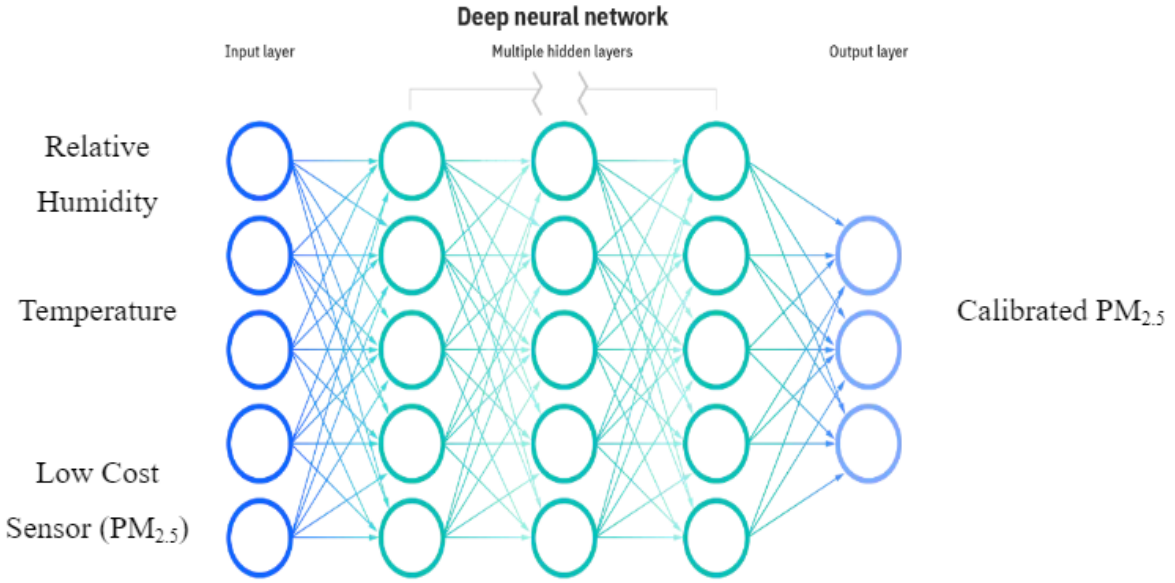


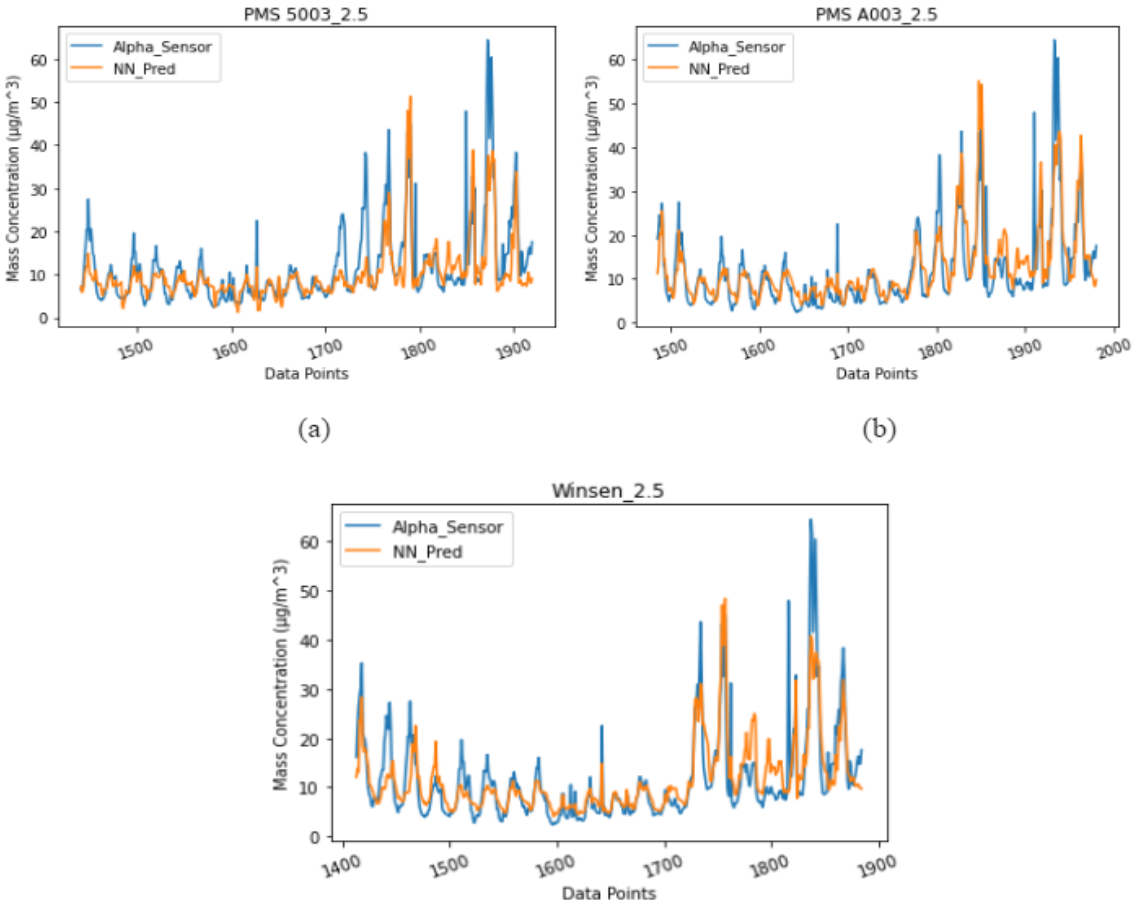
Figure 4.7 An illustration of a Artificial Neural Network with multiple hidden layers.

4.7.1 Results from Artificial Neural Network

The graph below compares the real concentration levels at the reference sensor (Alpha_sensor) with the obtained predicted values by low cost sensor (PM 5003) after calibration with Artificial Neural Network Algorithm

Table 4.20 Obtained results from artificial neural network algorithm when only relative humidity effect is considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.12	4.87	4.91
Mean Absolute Error (MAE)	3.54	3.09	3.17
Correlation	0.839	0.856	0.849



4.8 Relation between (a) PMS 5003 and alpha sensor, (b) PMS A003 and alpha sensor (c) Winsen and alpha sensor obtained from artificial neural network algorithm

Table 4.21 Obtained results from artificial neural network algorithm when only temperature effect is considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.12	4.87	4.91
Mean Absolute Error (MAE)	3.54	3.09	3.17
Correlation	0.839	0.856	0.849

Table 4.22 Obtained results from artificial neural network algorithm when only relative humidity effect is considered

Loss Function	PMS A003	PMS 5003	Winsen
Root Mean Square Error (RMSE)	5.12	4.87	4.91
Mean Absolute Error (MAE)	3.54	3.09	3.17
Correlation	0.839	0.856	0.849

Chapter 5: Characterization Chamber

5.1 Introduction

An acrylic sheet of 68×60×52 cm^3 dimension is used to make a characterization chamber for the particle measurement as we can see in the figure. Whose top side is detachable. The top portion of the chamber will be opened to enclose the sensors during the experiments, and then it will be closed with the sides sealed and the top portion replaced. Throughout the experiments, all open holes are sealed. For measurements, the sensors (PMS 5003, PMS 7003 and Winsen) will be put inside the chamber, while the reference sensors (Dusttrak) will be put outside and used to draw the aerosol sources with a tube. It is anticipated that particle loss in the tube will be negligible and that the tube length would be kept as short as possible. To run the sensor, we are using power banks. Initially no particles are inserted inside the characteristic chamber. Aerosol particles will be added after an hour, and the PM concentration will be monitored until it returns to normal.

Figure 5.1 Characterization Chamber in which experiment performed

To generate aerosol particles, we used 0.5% KCL Wt. solution. With the help of pneumatic pipes, these aerosol particles go to the characterization chamber, where sensors are mounted and measurements are performed. To dilute the concentration of aerosols, clean dry air will be inserted inside the aerosol chamber. For clean and dry air, we used a pump, dryer

and HEPA filter. Experiment is also carried out for humidity of 40% and 60% and then again aerosol particles are inserted inside the characterization chamber to analyze the effect on mass concentration and correlation.

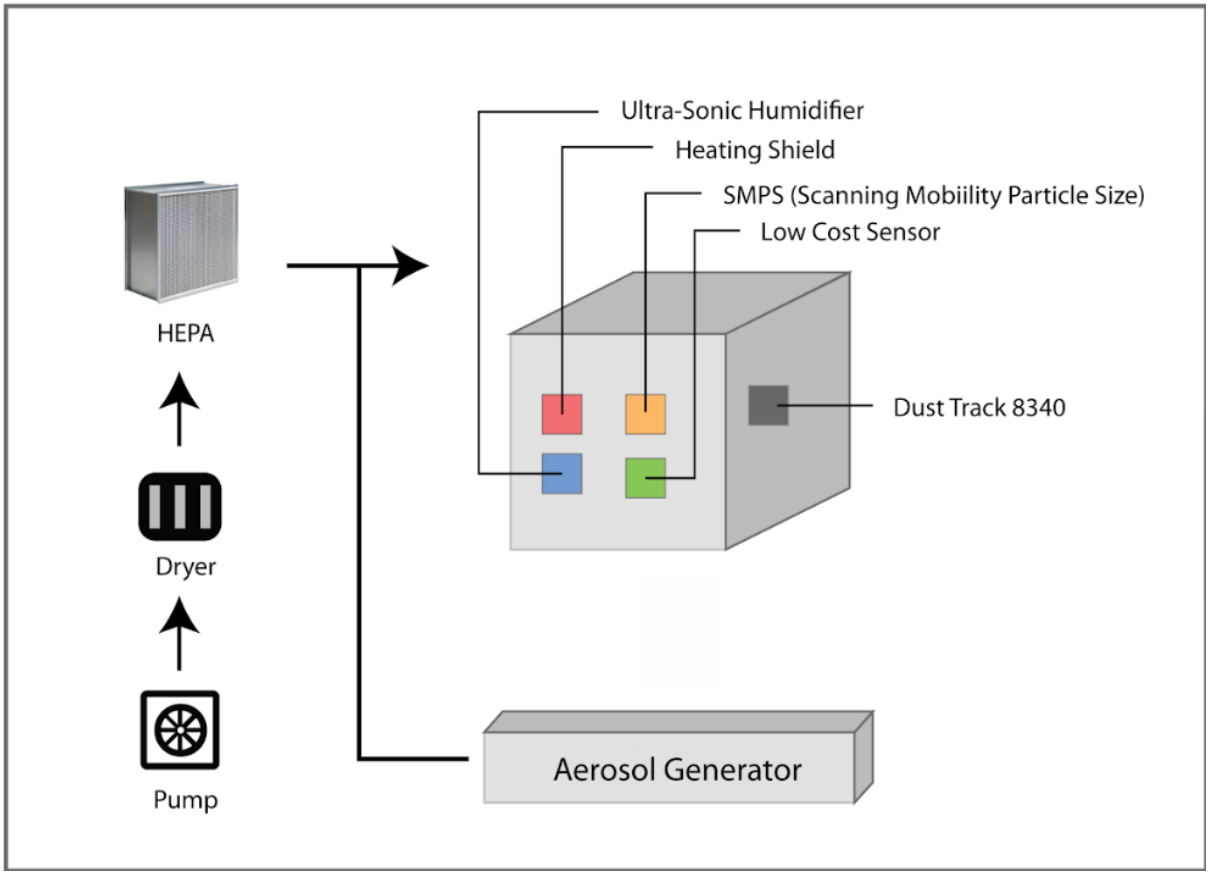


Figure 5.2 Characterization Chamber

Pump: Used to insert atmospheric air inside the Characterization Chamber

Dryer: It removes all the water particles from the atmospheric air that we are inserting and make the air dry

HEPA (High Efficiency Particulate Air): It works by forcing air through a fine mesh that traps harmful particles such as pollen, pet dander, dust mites, and tobacco smoke, so that we can regulate the PM particle present in the characterization chamber by generating and inserting PM particles from Aerosol Generator.

Ultrasonic Humidifier: It is used to reduce the humidity in the characterization chamber which helped us to perform experiment at different humidity

Heating Shield: Used to increase the temperature inside the characterization chamber which helped us to do experiment at different temperatures

SMPS (Scanning Mobility Particle Size): A scanning mobility particle sizer will be used to measure the distribution of particle numbers (SMPS Model number 3082). The SMPS

combines a condensation particle counter (CPC), which counts the fractionated particles, with a differential mobility analyzer (DMA), which divides particle sizes based on electrical mobility. Charged aerosol particles enter the DMA from the top. Higher-mobility particles hit the electrode before leaving, whereas lower-mobility particles discharge with more air. The particles that exited the sample outlet through the DMA are then directed into the CPC, which has a condenser stage and a saturator stage. In the heated saturator step, the particles pass through a saturated butanol vapor. In the cooled condenser, butanol vapor condenses on the particle surface, increasing particle size and making it simpler to measure particle size with laser light scattering techniques.

Dusttrak: It is used to measure the mass concentration of PM particles inside the chamber. It operates on the principle of light scattering using a laser diode that emits at a wavelength of 780 nm and it is built with a lens that is 90 degrees from the aerosol stream and the laser beam and concentrates the scattered light into a photodiode. The signal from the photodiode is converted into voltage which is directly proportional to the aerosol mass concentration.

5.2 Results:

5.2.1 Experiment performed at Humidity 40% and 60%

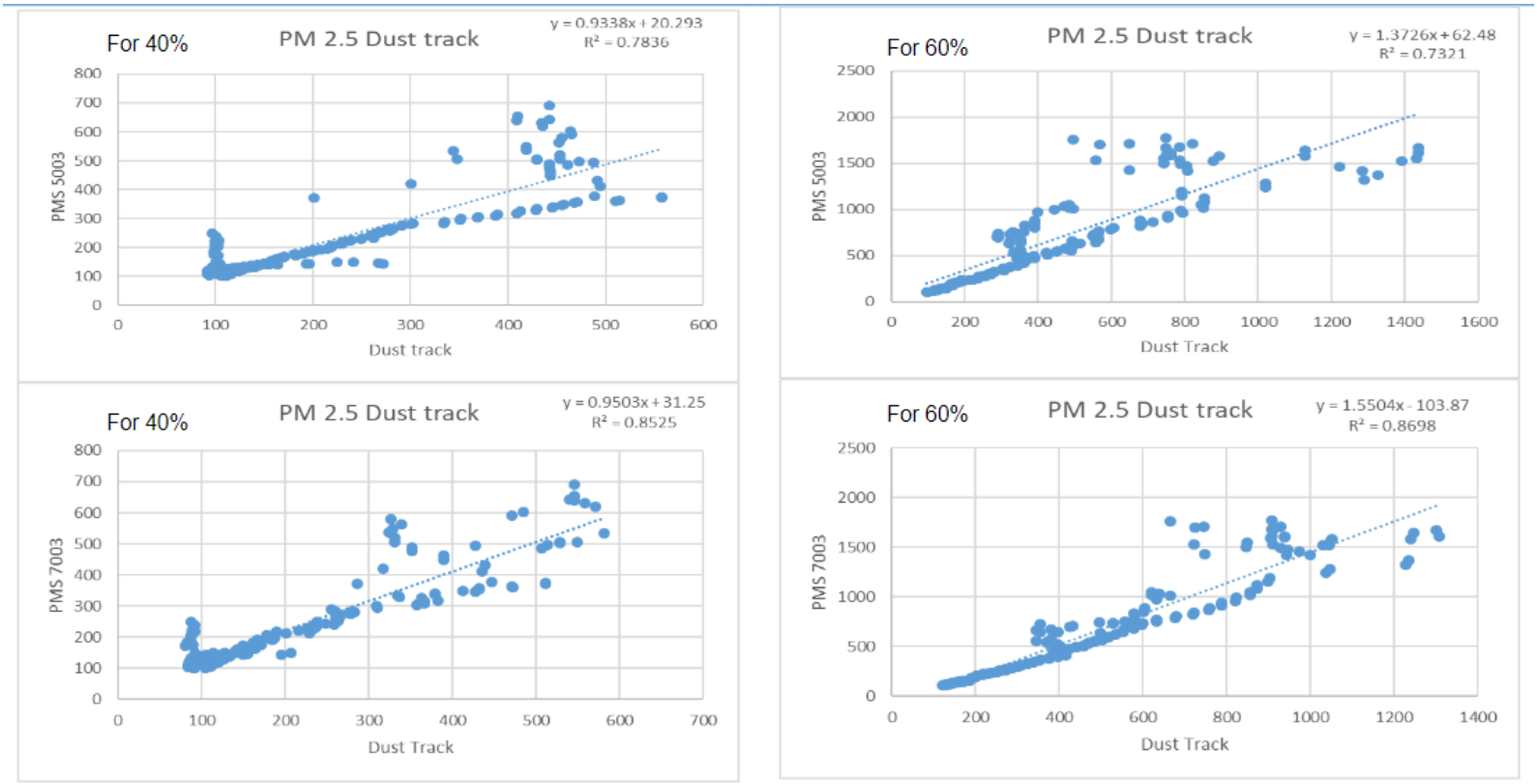


Figure 5.3 Correlation Obtained at different humidity

Table 5.1 Correlation Obtained by PMS 7003 and PMS 5003 at 40% and 60% Humidity

Correlation	PMS 7003	PMS 5003
40%	85.2	78.3
60%	86.9	73.2

From the Figure 5.3 and Table 5.1 we can conclude that:

- At lower mass concentration of PM particles, we are having close values but as the mass concentration or PM value increases we have seen the deviation.
- We are getting better results with PMS 7003 as compared to PMS 5003 in both the cases when we consider 40% and 60% humidity.

5.2.2 Variation of PM concentration at Humidity 40% and 60%

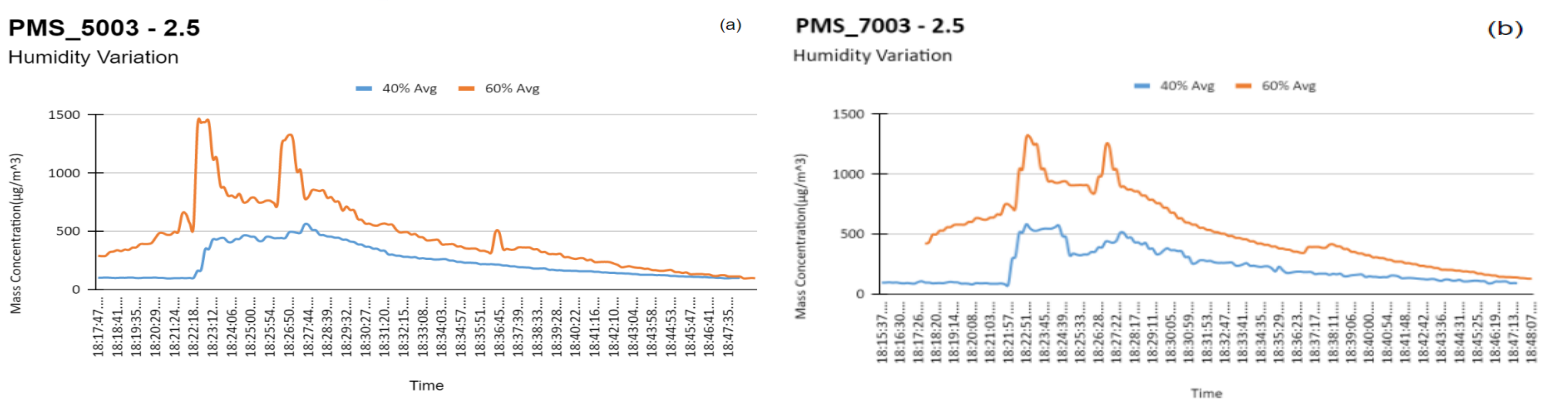


Figure 5.4 Variation of PM 2.5 Concentration at 40% and 60% Humidity (a) PMS_5003 (b) PMS_7003

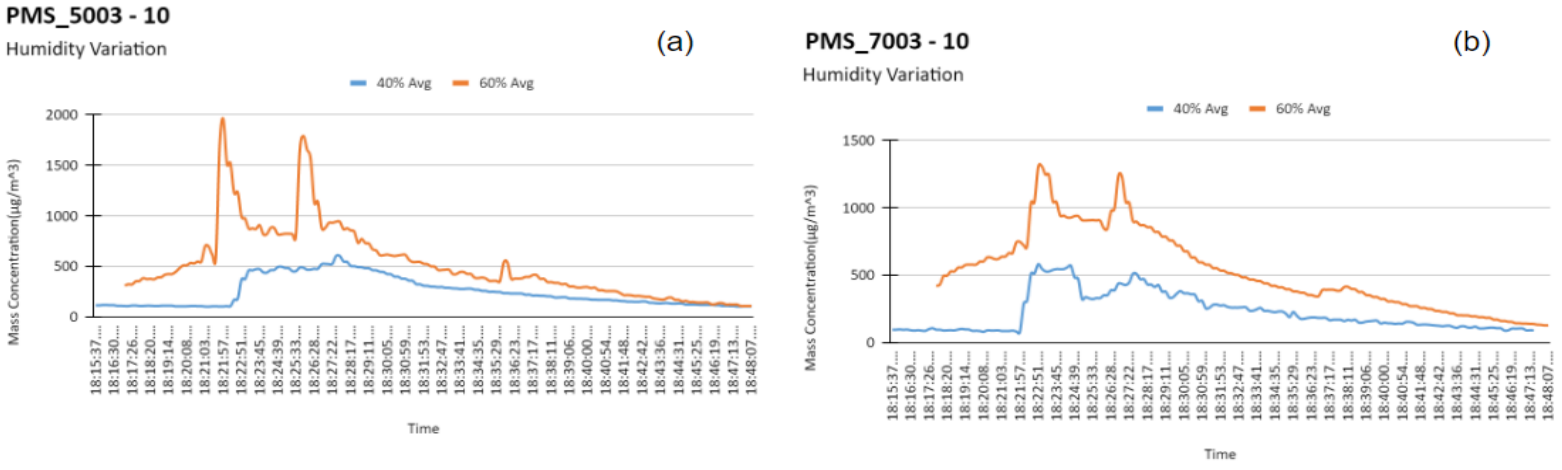


Figure 5.5 Variation of PM 10 Concentration at 40% and 60% Humidity (a) PMS_5003 (b) PMS_7003

From the Figure 5.4 and Figure 5.5 we can see that the Mass Concentration increases as we increase the humidity because when we increase the humidity concentration the amount of water vapor also increases in the characterization chamber which covers the upper surface of the Aerosol particle which results in increase in size of PM Particle.

5.2.3 Variation of PM concentration when aerosol inserted at Humidity 40% and 60%

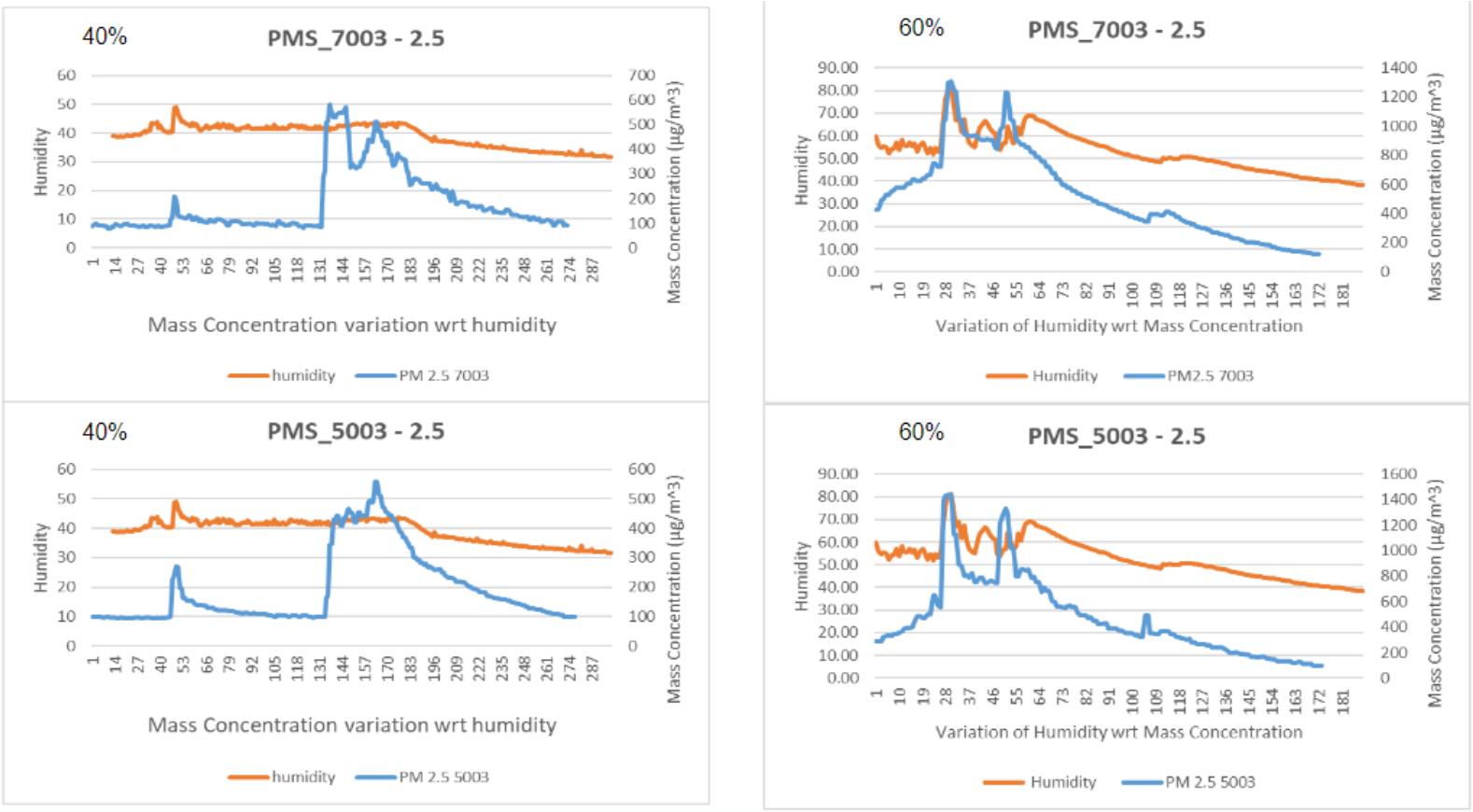


Figure 5.6 Variation of PM Concentration when aerosol inserted at 40% and 60% Humidity

Figure 5.6 for both the cases at humidity 40% and 60% for both PMS 5003 and PMS 7003 shows us the relationship between the Humidity and Mass Concentration of PM particles. Here initially we kept the humidity constant and then after some time we inserted aerosol particles which resulted in sudden increment in mass concentration and after that we gradually decreased the humidity which resulted in decrease of mass concentration of PM Particles.

5.2.4 Distribution of PM Particles inside the Characterization chamber when aerosol inserted (5% KCL Wt.)

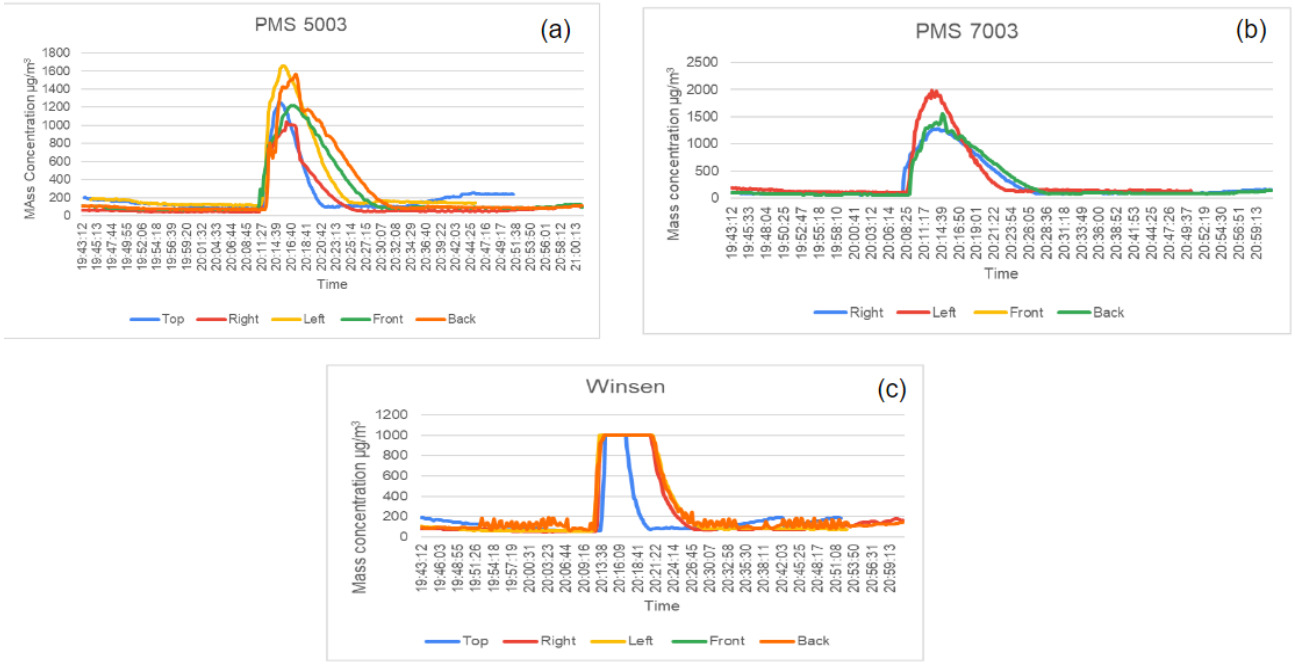


Figure 5.7 Variation of PM Concentration inside the characterization chamber when aerosol inserted at 40% and 60% Humidity on 27 Feb (a) PMS_5003 (b) PMS_7003 (c) Winsen

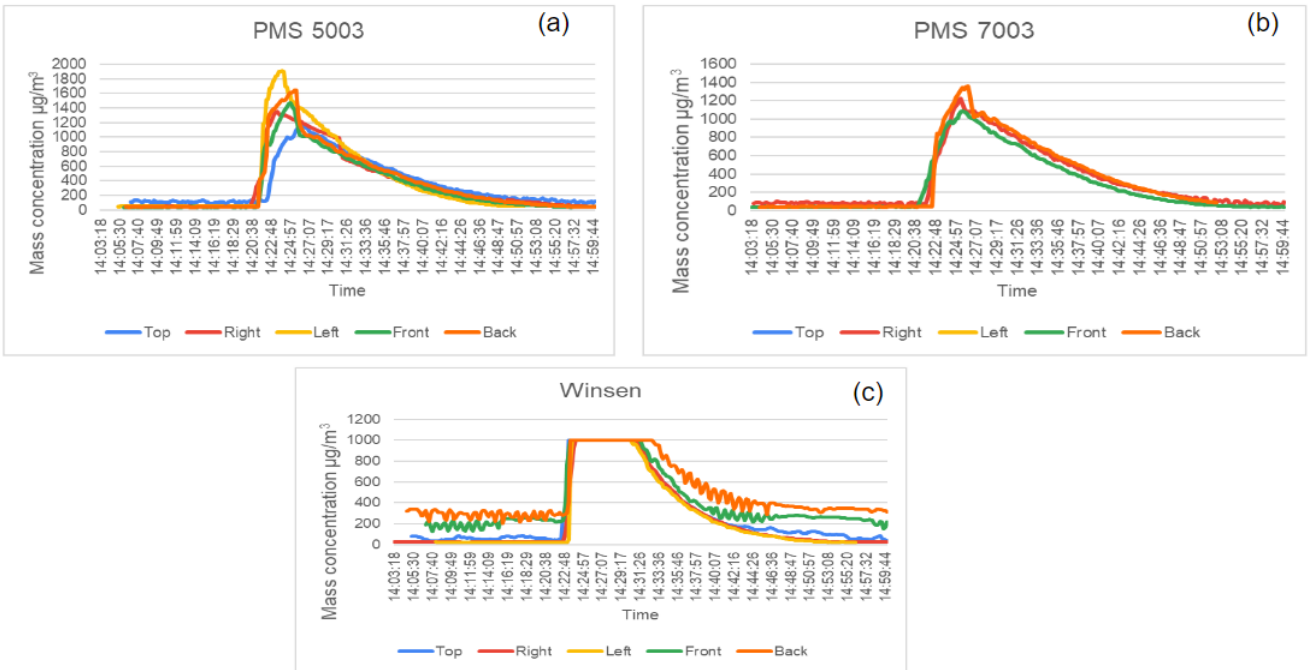


Figure 5.8 Variation of PM Concentration inside the characterization chamber when aerosol inserted at 40% and 60% Humidity on 26 Feb (a) PMS_5003 (b) PMS_7003 (c) Winsen

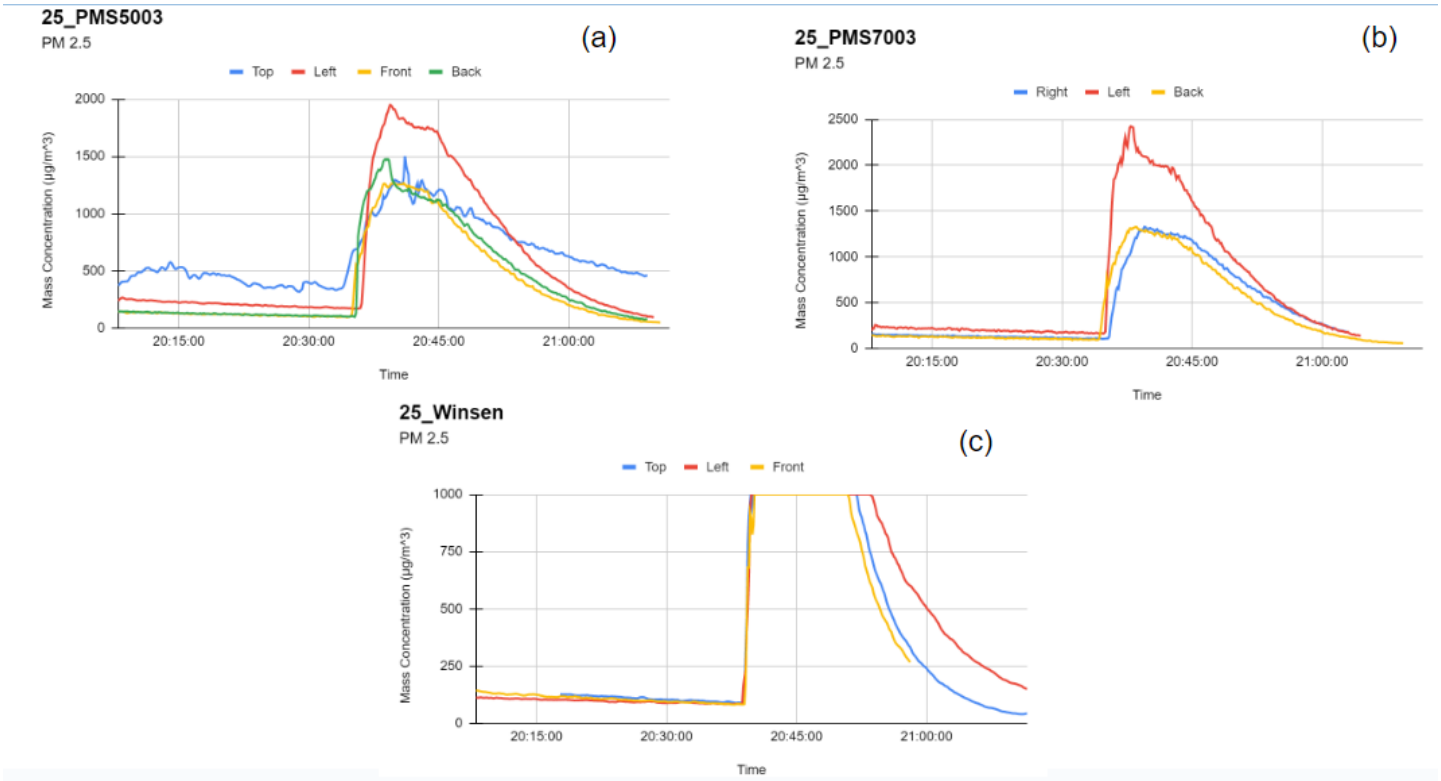


Figure 5.9 Variation of PM Concentration inside the characterization chamber when aerosol inserted at 40% and 60% Humidity on 25 Feb (a) PMS_5003 (b) PMS_7003 (c) Winsen

The purpose of this experiment was to analyze the distribution of aerosol or PM particles inside the characterization chamber. Therefore, in this experiment we have placed the sensor in the left, Right, Front, Back and Top (at some height) sides inside the characterization chamber. The experimental setup was divided into 3 stages. In first stage the chamber was cleaned by inserting diluted clean air while the aerosol valve was closed till the PM concentration become minimum, then on second stage the aerosol valve opened and then aerosol is inserted into the chamber in that we have seen in Figure 5.7, Figure 5.8 and Figure 5.9 the sudden increment in the PM concentration then when we reach the maximum value then in third stage again we stopped the aerosol valve and then we have seen the decay in PM concentration at each point inside the characterization chamber.

5.2.5 Correlation between Dust Particle and Low Cost Sensor

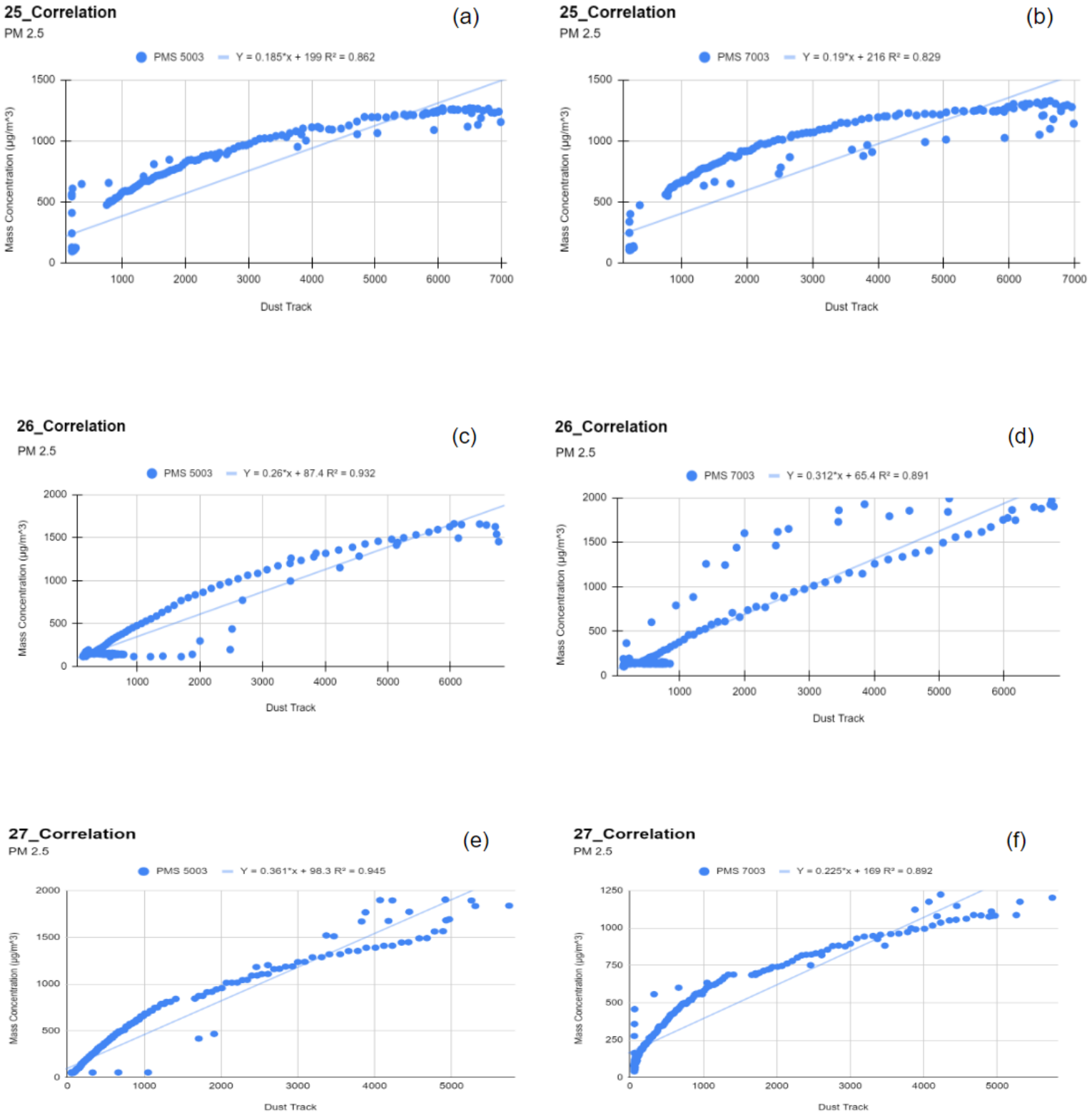


Figure 5.10 Correlation between the PM particle and Dusttrak

(a) PMS_5003 on 25th (b) PMS_7003 on 25th (c) PMS_5003 on 26th (d) PMS_7003 on 26th

(e) PMS_5003 on 27th (f) PMS_7003 on 27th

Table 5.2 Correlation Obtained by PMS 7003 and PMS 5003 at different location inside Characterization Chamber

	25					26					27				
Correlation	Top	Left	Right	Front	Back	Top	Left	Right	Front	Back	Top	Left	Right	Front	Back
PMS 5003	74.5	84.2	-	86.2	80.6	91.9	93.2	82.4	80.3	75.8	88.7	94.5	89.3	85.2	72.8
PMS 7003	-	85.7	82.9	-	82.5		94.6	84.1		78.9			90.2	89.2	86.6
Winsen						81.9	70.7	74		68.7	87.5	88.1	87.8	86	78.1

The purpose of the above experiment is to find how our low cost sensor placed inside the characterization chamber at different sides (Top, Left, Right, Front and Back) is correlated to the dust track (which is also a reference monitor). The first order polynomial regression gave a higher correlation value of 94.6 by PMS 7003 on 26th Feb inside the characterization chamber.

On 25th Feb we had installed and done testing so we are getting less correlation on that day but after that on further days we are getting better correlation as we can see in Table 5.2 and Figure 5.10.

Chapter 6: Results, Conclusion and Future Scope

6.1 Overall Results

As we have used all the above machine learning algorithms to find the best correlation between the low cost sensor (Low Cost Sensor) and reference sensor(Alpha_sensor) for the Particulate Matter of size 2.5μm and the result then we obtained are mentioned below for different sensors.

As low cost sensors are very sensitive towards environmental changes therefore we have seen the correlation when we have considered all the factors like Humidity, Temperature, Wind Speed and Wind Direction.

6.1.1 PMS 5003 results

Table 6.1 Correlations obtained from PMS 5003 when cumulative and individual effects of all the parameters are considered for different Machine learning algorithms

PMS 5003	Correlation when Temperature, Relative Humidity, Wind Speed and Wind Direction considered			
Algorithms	All Parameter Considered	Temperature & Humidity Considered	Only Temperature	Only Humidity
Linear Regression	0.67	0.67	0.67	0.67
Random Forest	0.853	0.833	0.73	0.729
Neural Network	0.832	0.856	0.77	0.777
K-Nearest Neighbor	0.846	0.817	0.76	0.757
Gaussian Process	0.851	0.848	0.758	0.75
Multiple Regression	0.847	0.849	0.757	0.756
Support Vector Method	0.85	0.846	0.754	0.756

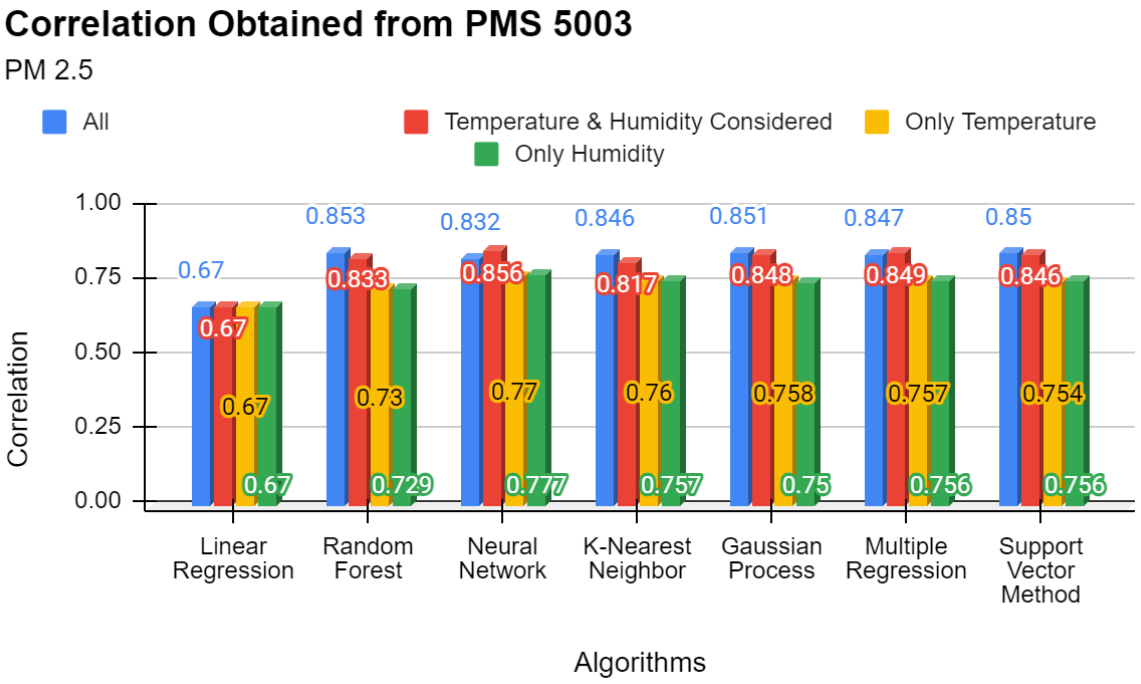


Figure 6.1 Correlations obtained from PMS 5003 when cumulative and individual effects of all the parameters are considered for different Machine learning algorithms.

From Table 6.1 and Figure 6.1 we can say that environmental factors strongly affect the correlation of low cost sensors as our correlation increases in most of the algorithms (like Random Forest, K-Nearest Neighbor, Gaussian Process, Support Vector Method algorithms) when we consider all the environmental factors for PMS 5003.

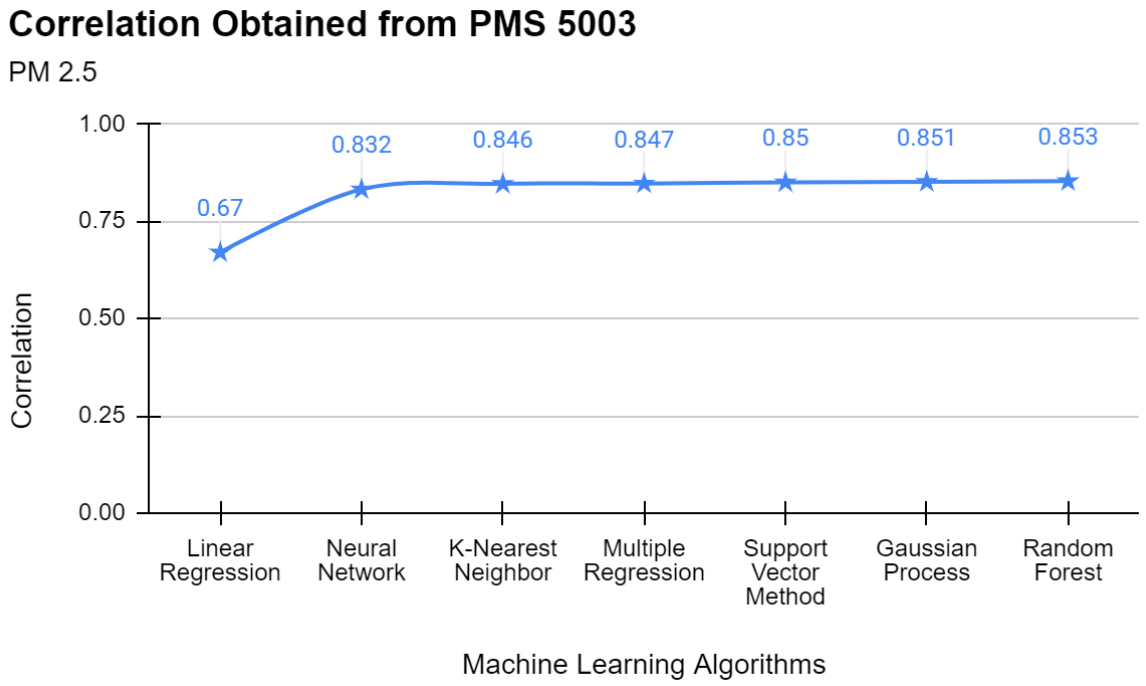


Figure 6.2 Correlation obtained from different Machine Learning Algorithms for PMS 5003

Figure 6.2 says that the machine learning algorithm (Random Forest, Support Vector Method and Gaussian progression algorithm) are providing better results

6.1.2 PMS A003 results

Table 6.2 Correlations obtained from PMS A003 when cumulative and individual effects of all the parameters are considered for different Machine learning algorithms

PMS A003	Correlation when Temperature, Relative Humidity, Wind Speed and Wind Direction considered			
Algorithms	All Parameter Considered	Temperature & Humidity Considered	Only Temperature	Only Humidity
Linear Regression	0.61	0.61	0.61	0.61
Random Forest	0.817	0.816	0.7	0.7
Neural Network	0.841	0.849	0.75	0.767
K-Nearest Neighbor	0.826	0.811	0.725	0.744
Gaussian Process	0.834	0.83	0.737	0.73

Multiple Regression	0.829	0.83	0.739	0.735
Support Vector Method	0.806	0.788	0.735	0.74

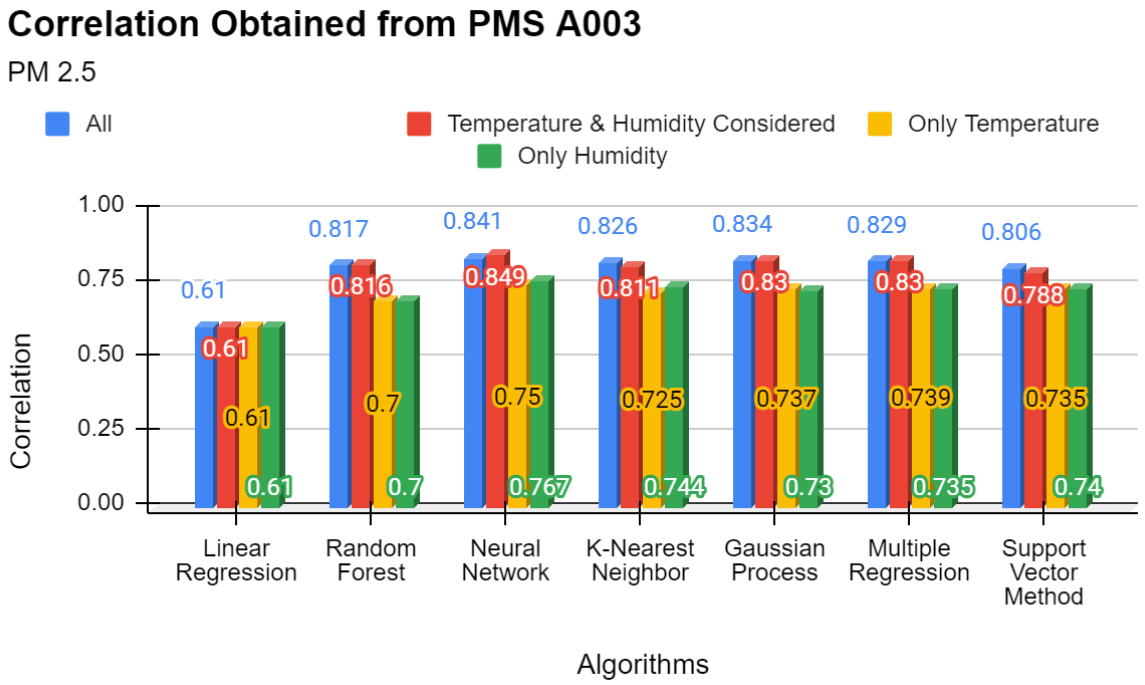


Figure 6.3 Correlations obtained from PMS A003 when cumulative and individual effects of all the parameters are considered for different Machine learning algorithms.

From Table 6.2 and Figure 6.3 we can say that environmental factors strongly affect the correlation of low cost sensors as our correlation increases in most of the algorithms (like Support Vector Method, K-Nearest Neighbor, Gaussian Process algorithm) when we consider all the environmental factors for PMS A003.

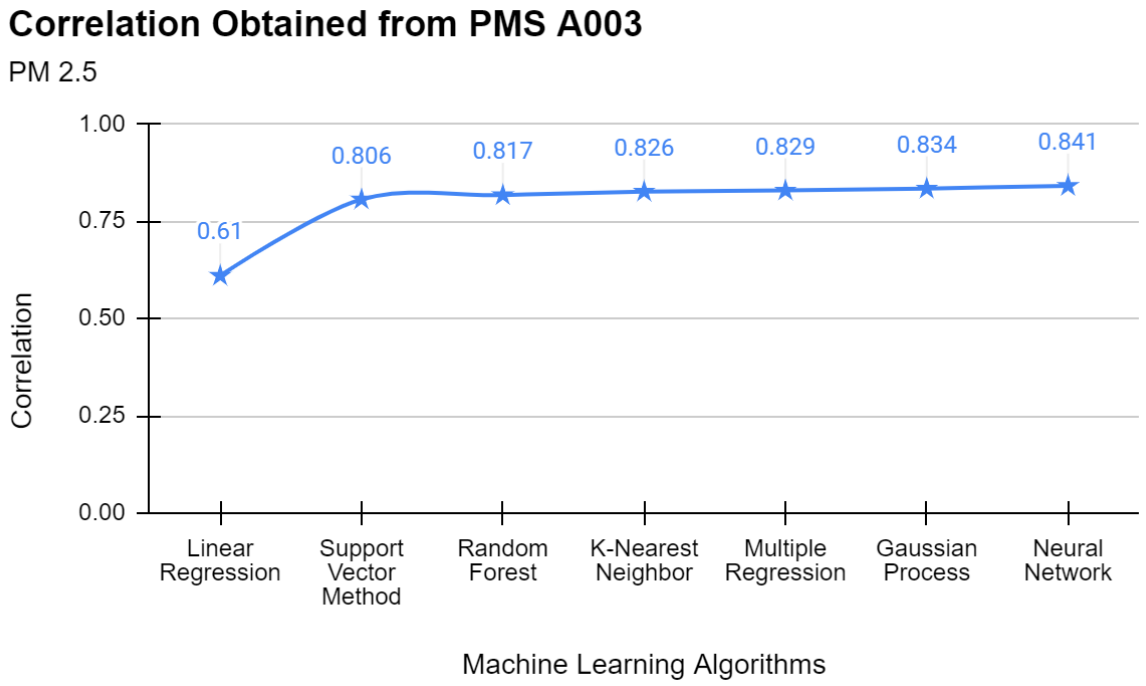


Figure 6.4 Correlation obtained from different Machine Learning Algorithms for PMS A003

Figure 6.4 says that the machine learning algorithms (artificial neural network, multiple linear regression and Gaussian Process) are providing better results for PMS A003.

6.1.3 Winsen results

Table 6.3 Correlations obtained from Winsen when cumulative and individual effects of both the sensors are considered for different Machine learning algorithms

Winsen	Correlation when Temperature, Relative Humidity, Wind Speed and Wind Direction considered			
Algorithms	All Parameter Considered	Temperature & Humidity Considered	Only Temperature	Only Humidity
Linear Regression	0.68	0.68	0.68	0.68
Random Forest	0.838	0.807	0.71	0.73
Neural Network	0.848	0.839	0.748	0.766
K-Nearest Neighbor	0.83	0.841	0.741	0.766
Gaussian Process	0.795	0.785	0.692	0.696

Multiple Regression	0.793	0.787	0.693	0.696
Support Vector Method	0.84	0.83	0.692	0.69

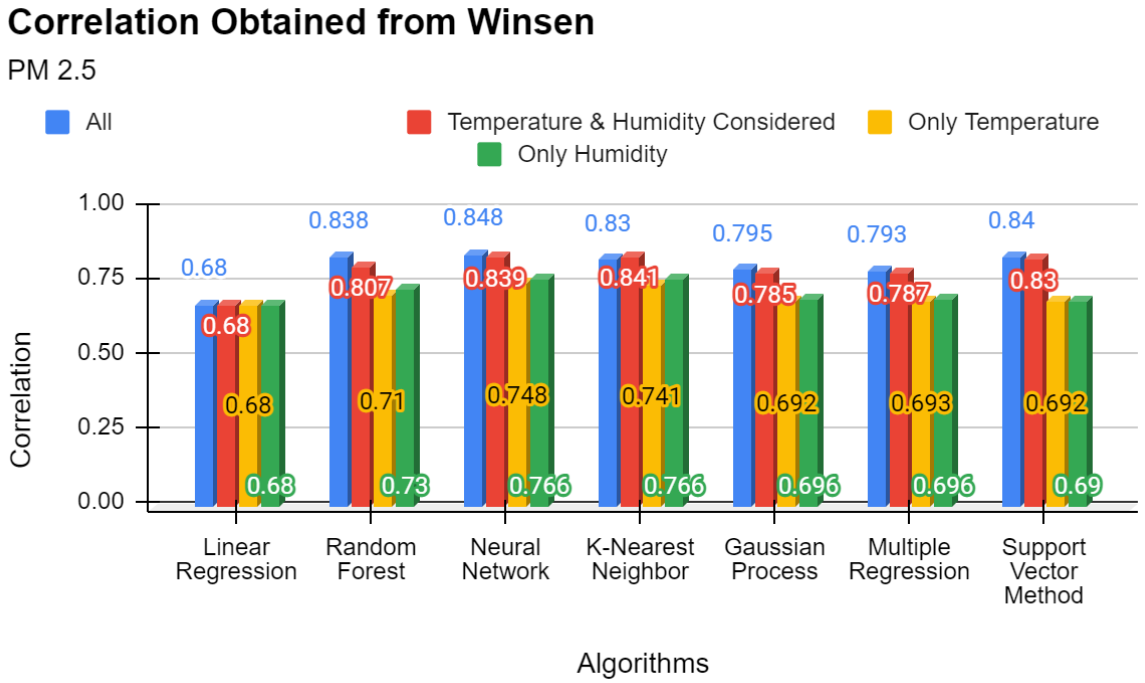


Figure 6.5 Correlations obtained from Winsen when cumulative and individual effects of both the sensors are considered for different Machine learning algorithms

From Table 6.3 and Figure 6.5 we can say that environmental factors strongly affect the correlation of low cost sensors as our correlation increases in most of the algorithms like Support Vector Method, Multiple Linear Regression, Gaussian Process algorithms when we consider all the environmental factors for Winsen.

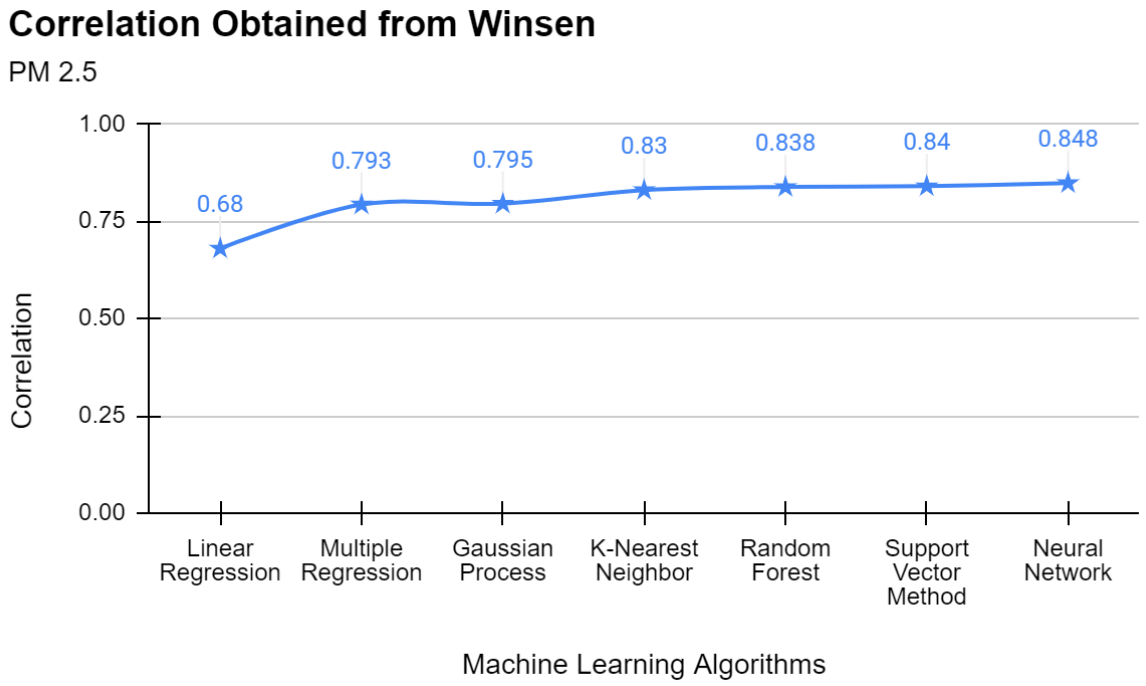


Figure 6.6 Correlation obtained from different Machine Learning Algorithms for Winsen

Figure 6.6 says that the machine learning algorithms (artificial neural network, Support Vector Method and Random Forest) are providing better results for Winsen.

6.1.4 Low cost sensor comparison

Table 6.4 Comparison of correlation between different low cost sensors

Algorithms	PMS 5003	PMS A003	Winsen
Linear Regression	0.67	0.61	0.68
Random Forest	0.853	0.817	0.838
Neural Network	0.832	0.841	0.848
K-Nearest Neighbor	0.846	0.826	0.83
Gaussian Process	0.851	0.834	0.795
Multiple Regression	0.847	0.829	0.793
Support Vector Method	0.85	0.806	0.84

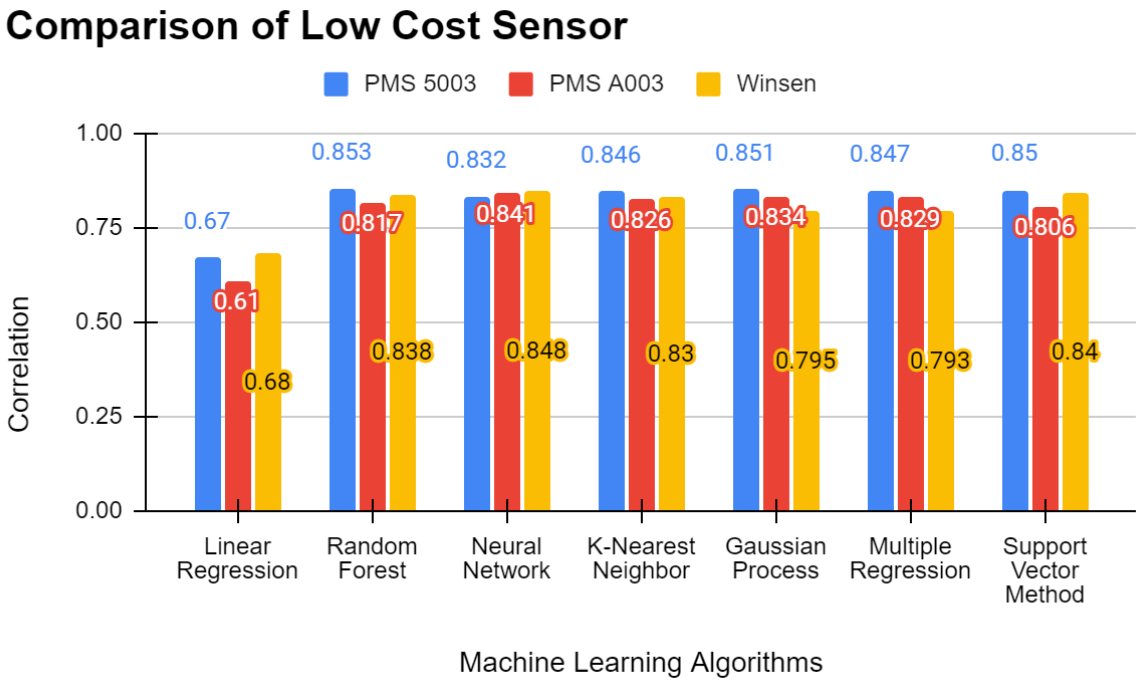


Figure 6.7 Comparison of correlation between different low cost sensors

From Table 6.4 and Figure 6.7 we can say that all the low cost sensors prediction values are very close to each other, however the best results we are getting are from PMS 5003 where the maximum value of correlation is 0.853 obtained by Random Forest algorithm.

Different low cost sensors have different accuracy levels therefore by comparing all the three low cost sensors (PMS A003, PMS 5003 and Winsen) from the above figure 6.7. As per their performance I have given the rank for low cost sensor.

Table 6.5 Ranking of sensor according to their prediction performance

Rank	Low Cost Sensor
1	PMS 5003
2	Winsen
3	PMS A003

6.1.5 Seasonality

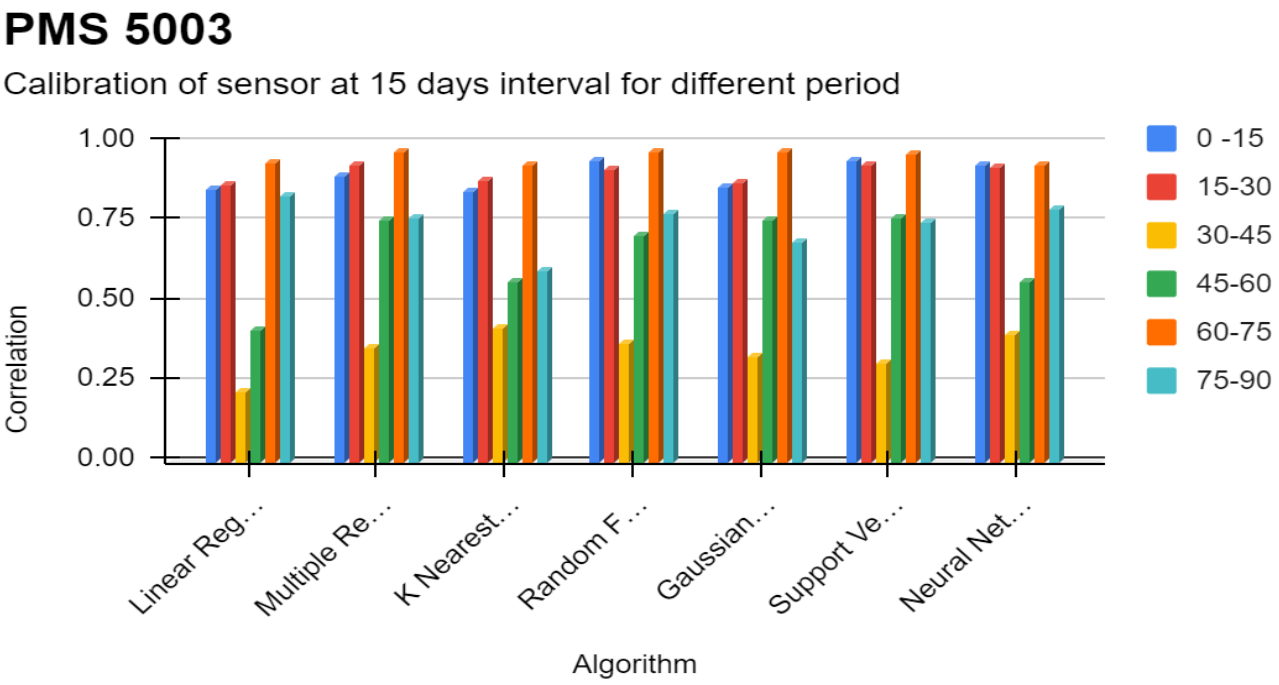


Figure 6.8 Correlation obtained for 15 days interval period

As we know, seasonality plays significant role for change in correlation similar thing we have seen when we have trained our data in Machine Learning algorithm there we have seen drop in correlation between alpha sensor which is our reference monitor and low cost sensor which encourages us to go find where the season started to change, for that we have distributed our dataset in the set of 15 days and then we trained the data against the machine learning algorithm and then we can see from the above Figure 6.8 the yellow block which shows us the period of 30-45 days having low correlation value for all algorithm which tells us that the seasonality change occur in that period.

6.1.6 Colocation Period

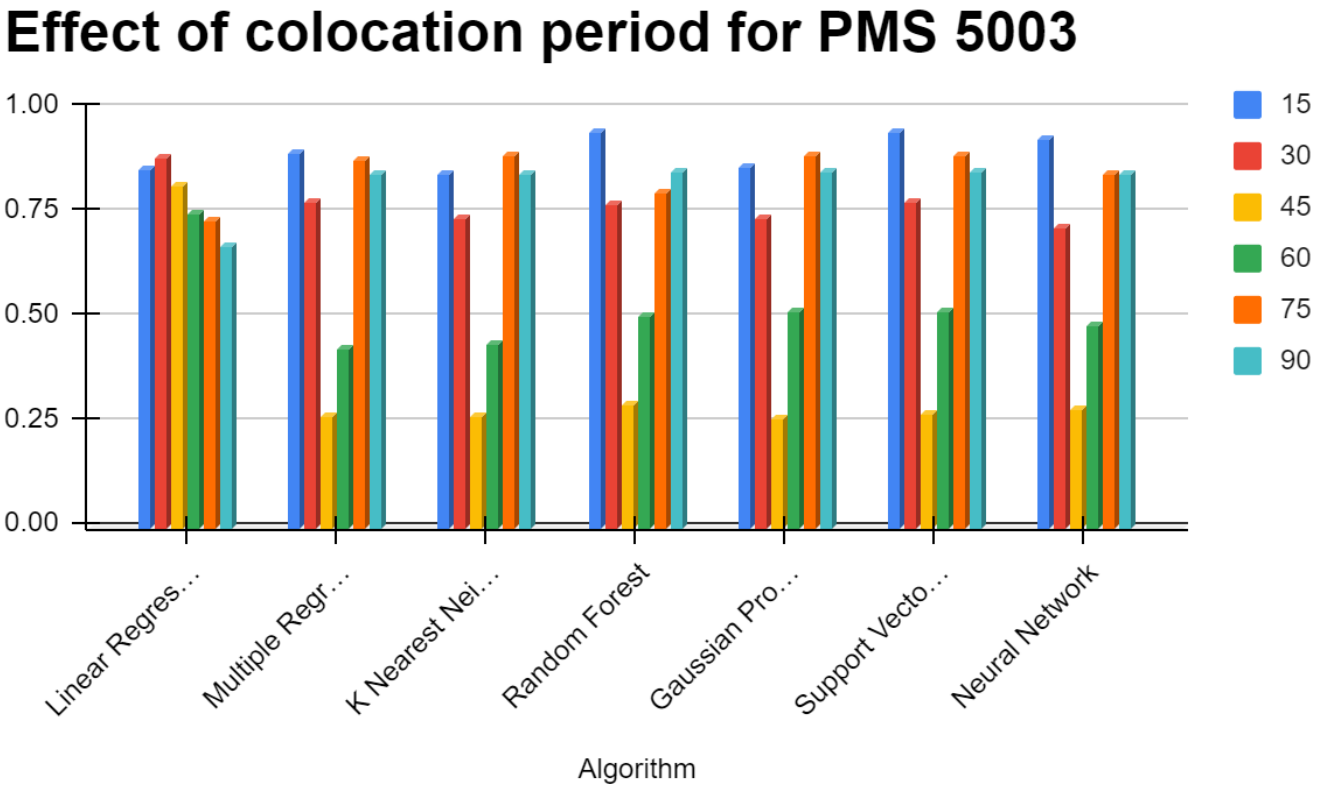


Figure 6.9 Correlation obtained when period increased by 15 days period

The purpose of this is to determine the sensitivity of length of collocation period and to see the effect of correlation with different colocation periods for different machine learning algorithms. For that we have gradually increased the dataset by 15 days every time.

As we can see in the above graph the first block is for 15 days then next for 30 days and so on till 90 days. From the graph one interesting thing we have noticed is that in between we have seen a drop in correlation that is because of seasonality which we have already discussed above.

6.2 Conclusion

The correlations obtained for different low cost sensors (PMS A003, PMS 5003 and Winsen) when cumulative and individual effects are considered, from that we conclude the correlation increases when all the parameters are considered as input for different Machine learning algorithms.

Best Pearson correlation coefficients obtained from random forest, gaussian process and support vector method machine learning algorithms.

For calibration of low cost sensor, we have used various machine learning algorithm, out of which top 3 algorithms from which we are getting better performance are:

Table 6.6 Ranking of Machine Learning Algorithm according to their prediction performance measured in PMS 5003

Rank	Algorithm	Correlation	RMSE	MAE
1	Random Forest	0.853	4.87	3.09
2	Gaussian Process	0.851	17.53	13.9
3	Support Vector Method	0.85	4.93	3.08

While analyzing the data we have also seen the seasonality in the period of 30 - 45 days in Figure 6.8 which highly impacted our correlation.

The characterization of an aerosol chamber for the assessment of PM sensors has been covered in this report. The performance of low cost sensors in the characterization chamber has increased to 0.94(Pearson correlation) by comparing it to Dusttrak 8533. At lower PM concentrations the value of Low Cost Sensor and Reference monitor are nearly close to each other but at higher PM concentrations the value of Low Cost Sensor deviates with respect to reference monitor. PM concentrations are also monitored for different humidity levels by inserting aerosol particles, for high humidity PM concentrations is high and for lower humidity PM concentration is low. In the characterization chamber PMS7003 is showing better correlation as compared to PMS5003.

6.3 Future Work

- To mitigate the effect of relative humidity, an air drying system or heater will also be used at the inlet of low cost sensors, to analyze the correlation with respect to the different algorithms for low cost sensors like PMS5003, Winsen, SPS30 and honeywell.
- The calibrated sensors will be used in the characterization chamber to ensure the better correlation between LCS and reference monitor.
- The Calibration of LCS will be done by using different aerosol sources like incense sticks, candle burning, etc.

References

[1] Fadhli, M., Asriyadi, A., Lindawati, L., Salamah, I., Affrylia, G., Valerie, M., and Ramadhan, A. (2022). Low Cost Air Quality Monitoring System Using LoRa Communication Technology. Proc. 5th FIRST T1 T2 2021 Int. Conf. (FIRST-T1-T2 2021) 9:297–302. doi:10.2991/ahe.k.220205.052.

[2] Kan, H., Chen, R., and Tong, S. (2012). Ambient air pollution, climate change, and population health in China. Environ. Int. (2012) doi:10.1016/j.envint.2011.03.003

[3] Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., Bell, M., Norford, L., and Britter, R. (2015). The rise of low-cost sensing for managing air pollution in cities. Environ. Int. 75:199–205. doi:10.1016/j.envint.2014.11.019.

[4] Kuula, J., Timonen, H., Niemi, J. V., Manninen, H.E., Rönkkö, T., Hussein, T., Fung, P.L., Tarkoma, S., Laakso, M., Saukko, E., Ovaska, A., Kulmala, M., Karppinen, A., Johansson, L., and Petäjä, T. (2022). Opinion: Insights into updating Ambient Air Quality Directive 2008/50/EC. Atmos. Chem. Phys. 22 (7):4801–4808. doi:10.5194/acp22-4801-2022.

[5] Li, J. and Williams, B. (2019). Recent advances in low-cost particulate matter sensor : calibration and application.

[6] Liyao Yang, Cheng Li and Xiaoxiao Tang (2020). The Impact of PM2.5 on the Host Defense of Respiratory System. https://doi.org/10.3389/fcell.2020.00091

[7] Lung, S.C.C., Hien, T.T., Cambaliza, M.O.L., Hlaing, O.M.T., Oanh, N.T.K., Latif, M.T., Lestari, P., Salam, A., Lee, S.Y., Wang, W.C.V., Tsou, M.C.M., Cong-Thanh, T., Cruz, M.T., Tantrakarnapa, K., Othman, M., Roy, S., Dang, T.N., and Agustian, D. (2022). Research Priorities of Applying Low-Cost PM2.5 Sensors in Southeast Asian Countries. Int. J. Environ. Res. Public Health 19 (3). doi:10.3390/ijerph19031522.

[8] Mendez, E.; Temby, O.; Wladyka, D.; Sepielak, K.; Raysoni, A.U. Using Low-Cost Sensors to Assess PM2.5 Concentrations at Four South Texan Cities on the U.S.—Mexico Border. Atmosphere 2022, 13, 1554. https://doi.org/ 10.3390/atmos13101554

[9] Michael R.Giordanoab, Carl Malingsc, Spyros N.Pandisde, Albert A.Prest, V.F.McNeill, Daniel M.Westervelt, MatthiasBeekmannaj, R.Subramanian. From low-cost sensors to high-quality data: A summary of challenges and best practices for

effectively calibrating low-cost particulate matter mass sensors. https://doi.org/10.1016/j.jaerosci.2021.105833

[10] Minxing Si1,2,, Ying Xiong1,, Shan Du3 and Ke Du1 (2020). Evaluation and calibration of a low-cost particle sensor in ambient conditions using machine-learning methods. https://doi.org/10.5194/amt-13-1693-2020

[11] Najeeb Ullah, Faizullah Khan, Abdul Ali Khan, Surat Khan, Abdul Wahid Tareen, Muhammad Saeed, Akbar Khan. Optimal real-time static and dynamic air quality monitoring system. Indian Journal of Science and Technology. 2020; 13 (01),91-102. DOI: 10.17485/ijst/2020/v013i01/148375

[12] Omidvarborna, H., Kumar, P., and Tiwari, A. (2020). 'EnvilutionTM' chamber for 43 performance evaluation of low-cost sensors. Atmos. Environ. 223:117264. doi:10.1016/j.atmosenv.2020.117264.

[13] Park, D.; Yoo, G.-W.; Park, S.-H.; Lee, J.-H. Assessment and Calibration of a Low-Cost PM2.5 Sensor Using Machine Learning (HybridLSTM Neural Network): Feasibility Study to Build an Air Quality Monitoring System. Atmosphere 2021, 12, 1306. https://doi.org/10.3390/atmos12101306

[14] Wang, Y., Li, J., Jing, H., Zhang, Q., Jiang, J., and Biswas, P. (2015). Laboratory Evaluation and Calibration of Three Low-Cost Particle Sensors for Particulate Matter Measurement. Aerosol Sci. Technol. 49 (11):1063–1077. doi:10.1080/02786826.2015.1100710

[15] Yoo, G.-W.; Park, S.-H.; Lee, J.-H. Assessment and Calibration of a Low-Cost PM2.5 Sensor Using Machine Learning (HybridLSTM Neural Network): Feasibility Study to Build an Air Quality Monitoring System. Atmosphere 2021, 12, 1306. https://doi.org/10.3390/atmos12101306

[16] Aerosol and Air Quality Research, 19: 181–194, 2019 Copyright © Taiwan Association for Aerosol Research ISSN: 1680-8584 print / 2071-1409 online doi: 10.4209/aaqr.2017.12.0611

[17] Li, Jiayu, "Recent advances in low-cost particulate matter sensor: calibration and application" (2019). McKelvey School of Engineering Theses & Dissertations. 450. https://openscholarship.wustl.edu/eng_etds/450

[18] Chatzidiakou, L., Krause, A., Popoola, O., Di Antonio, A., Kellaway, M., Han, Y., Squires, F., Wang, T., Zhang, H., Wang, Q., Fan, Y., Chen, S., Hu, M., Quint, J., Barratt, B., Kelly, F., Zhu, T., and Jones, R. (2019). Characterising low-cost sensors in

highly portable platforms to quantify personal exposure in diverse environments. Atmos. Meas. Tech. 12 (8):4643–4657. doi:10.5194/amt-12-4643-2019.

[19] Moreno-Rangel, A., Sharpe, T., Musau, F., and McGill, G. (2018). Field evaluation of a low cost indoor air quality monitor to quantify exposure to pollutants in residential environments. J. Sensors Sens. Syst. 7 (1):373–388. doi:10.5194/jsss-7-373-2018

[20] Austin, E., Novosselov, I., Seto, E., and Yost, M.G. (2015). Laboratory evaluation of the Shinyei PPD42NS low-cost particulate matter sensor. PLoS One 10 (9):1–17. doi:10.1371/journal.pone.0137789